

The genetic basis for response to the Ketogenic diet in drug- resistant epilepsy

Natasha Emma Schoeler

A thesis for submission to UCL for the degree of

Doctor of Philosophy

Declaration

I, Natasha Emma Schoeler, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signed:

.....

Date:

Statement of contribution

The idea for this thesis was a joint effort between Profs Sanjay Sisodiya, Helen Cross, Ley Sander and I (I am referred to as 'the researcher' throughout the thesis).

I completed all ethics applications and amendments.

All recruitment and collection of blood samples was undertaken by me, with the exception of participants recruited from Bristol Royal Hospital for Sick Children, Alder Hey Children's Hospital and The Royal Children's Hospital in Melbourne; some participants from Birmingham Children's Hospital, St George's Hospital and Matthew's Friends clinics were recruited by a specialist nurse (Bernie Concannon), a dietitian (Orla Stone) or keto-assistant (Valerie Aldridge) respectively, and some participants were recruited by me.

I collected all phenotypic data, with the exception of individuals recruited from The Royal Children's Hospital in Melbourne, for whom Miss Jacinta McMahon provided phenotypic data.

DNA collected in the UK was extracted by clinical geneticists at GOSH.

I prepared all DNA samples to be sent to various centres for genotyping or sequencing.

Miss Anna Tostevin introduced me to the laboratory setting and assisted with some dilutions and packaging. Sequencing of *SLC2A1* was completed by Dr Suzanne Drury; sequencing of *KCNJ11* and *BAD* was completed by members of Professor Sian Ellard's team at Royal Devon & Exeter Hospital; whole exome sequencing was completed by Miss Deborah Hughes and Dr Alan Pittman.

Raw genotyping data from the Illumina HumanOmniExpressExome Beadchip was quality-control filtered and exported into PLINK format by Dr Costin Leu. The Reference Sequence Database file provided by PLINK/SEQ, used in exome sequencing analyses, was amended by

Dr Costin Leu. The effective number of independent tests and the significance threshold for the genome-wide association study was calculated by Dr Jon White. I performed all other statistical and genetic analyses.

Abstract

The Ketogenic diet (KD) is an alternative treatment option for people with drug-resistant epilepsy. It can reduce seizure frequency, but it is resource-intensive and may cause adverse side effects. Predictors of response – which, in the absence of specific metabolic disorders, are unknown – would improve patient selection and may enhance understanding of how the KD exerts its antiseizure effect. This project is concerned with identifying possible genetic markers of response to the KD. DNA was extracted from capillary blood taken from individuals who were following the KD for their epilepsy, or who had done so in the past. Individuals were classed as responders if they achieved ≥ 50 seizure reduction. Response was classified at various follow-up points, as well as a summary of response over time. Association studies were conducted using candidate gene (*KCNJ11* and *BAD*) sequencing, genome-wide single nucleotide polymorphism array, and whole exome sequencing data to determine whether there was an over-representation of specific gene variants in KD responders, compared to non-responders.

Common variation in *KCNJ11* and *BAD* was not significantly associated with KD response. rs12204701 reached significance in the array-based genome-wide association study including common variants, with 3-month diet response as the phenotype. No significant results were obtained when summary diet response was taken as the phenotype. Using the gene-based c-alpha test with the exome sequencing data, including all exonic and splicing variants, *ANKRD36C* reached significance; using a pathway-based count of case-unique alleles test, the 'ERBB1 Internalisation' pathway reached significance. No further significant results were obtained from the exome sequencing data when using other gene- and pathway-based tests or when variants were further filtered according to predicted functional consequence. Other genes with large differences in responder/non-responder minor or alternative allele counts are also of interest. It is unknown how these may

contribute to variability in KD response. Some common themes were identified amongst the genes and pathways of significance and suggestive significance: cell cycling, apoptosis, glucose homeostasis, neurological processes and triglyceride biosynthesis. It is biologically plausible that these processes influence KD response, although it is likely that many genes play a role.

A larger sample size is needed in order to improve power to detect genotypic-phenotypic associations and increase confidence in the importance of the genes of interest.

Table of Contents

Declaration	2
Statement of contribution.....	3
Abstract	5
List of Tables	14
List of Figures	16
List of abbreviations	19
Acknowledgements.....	22
1 Background: Epilepsy, the Ketogenic diet and Genetics	23
1.1 Epilepsy	23
1.1.1 Definitions	23
1.1.2 Burden	24
1.1.3 Treatment and management.....	25
1.2 The Ketogenic diet and Epilepsy.....	27
1.2.1 The Ketogenic diet – clinical reports of effectiveness.....	30
1.2.2 Effects of the Ketogenic diet beyond seizure frequency.....	40
1.2.3 Tolerability.....	42
1.2.4 Retention	43
1.2.5 Biochemical response to the Ketogenic diet and putative mechanisms of action	46
1.2.6 Are we able to predict response to the Ketogenic diet?.....	66
1.3 The Ketogenic diet and Genetics	78
1.3.1 Differential response to the Ketogenic diet	79
1.3.2 Differential regulation of gene expression	83
1.4 Research questions and hypotheses	92
2 Methodology.....	95
2.1 Ethics and recruitment.....	95

2.2	Phenotypic data collection	99
2.3	Genotypic data collection	101
2.3.1	Research questions 1 and 2	101
2.3.2	Research question 3	104
3	Phenotypic results	107
3.1	Introduction	107
3.2	Methods	109
3.3	Results	112
3.3.1	An overview	112
3.3.2	Compliance	117
3.3.3	Ketogenic diet response: seizure frequency at specific follow-up points	119
3.3.4	Factors beyond seizure frequency	123
3.3.5	Ketogenic diet response over time	125
3.3.6	Factors affecting Ketogenic diet response	128
3.3.6.1	Demographic and clinical parameters	128
3.3.6.2	Epilepsy syndrome or cause	133
3.3.6.3	Biochemical parameters	134
3.3.7	Use of phenotypic data for genetic analyses	136
3.4	Discussion	138
3.4.1	Cohort demographics	139
3.4.2	Compliance	140
3.4.3	Diet response	143
3.4.4	Factors affecting Ketogenic diet response	148
3.4.5	Factors beyond seizure frequency	149
3.4.6	Conclusion	151
4	Candidate gene analysis	153
4.1	Introduction	153

4.1.1	The candidate gene approach	153
4.1.2	Candidate genes: background information	156
4.1.3	Rationale for selection of candidate genes	160
4.2	Methods	163
4.2.1	Phenotypic data	163
4.2.2	Genetic data	165
4.2.2.1	Quality control	167
4.2.2.2	Association analysis	167
4.2.2.3	Population structure	169
4.2.2.4	Information on variants found	170
4.2.3	Power calculations	171
4.3	Results	172
4.3.1	Quality control	172
4.3.2	Common variation in <i>KCNJ11</i> and <i>BAD</i>	172
4.3.2.1	Overview	172
4.3.2.2	Population genetics	178
4.3.3	Association analysis	180
4.3.4	Previously unreported variants or SNPs with minor allele frequency <0.01	182
4.3.5	Haplotype analyses	187
4.3.6	Power calculations	187
4.4	Discussion	189
4.4.1	Potential consequences of SNPs with lowest p-values	192
4.4.2	Rare and novel variation in <i>KCNJ11</i> and <i>BAD</i>	196
4.4.3	Conclusions and future work	196
5	Common genetic variation and response to the Ketogenic diet: a Genome-wide Association Study	198
5.1	Introduction	198
5.1.1	The genome-wide association study approach	198

5.1.2	Quality control	199
5.1.3	Population stratification	201
5.1.4	The multiple testing burden	202
5.1.5	Interpreting results	203
5.1.6	Alternative approaches to the single-SNP method.....	204
5.2	Methods	209
5.2.1	Phenotypic data	209
5.2.2	Genetic data	210
5.2.3	Quality control	213
5.2.3.1	Per-individual quality control exclusion criteria	213
5.2.3.2	Per-SNP quality control exclusion criteria	215
5.2.4	Covariates and population structure	216
5.2.5	GWAS.....	221
5.2.6	Correction for multiple testing	221
5.2.7	Visualisation of GWAS results.....	222
5.2.8	Power calculations	224
5.2.9	Gene- and pathway-based analyses.....	224
5.3	Results.....	226
5.3.1	Quality control	226
5.3.2	Covariates and population structure	227
5.3.3	Correction for multiple testing	229
5.3.4	GWAS results	229
5.3.5	Power calculations	234
5.3.6	Gene- and pathway- based analyses.....	236
5.4	Discussion.....	246
5.4.1	Significant results	246
5.4.2	Potential sources of error	249

5.4.2.1	Phenotypic.....	249
5.4.2.2	Genotypic.....	250
5.4.3	SNPs of suggestive significance.....	252
5.4.4	Gene- and pathway-based analyses.....	254
5.4.5	Conclusions and future work.....	259
6	Rare genetic variation and response to the Ketogenic diet: Illumina exome chip	261
6.1	Introduction.....	261
6.1.1	The common disease/rare variant hypothesis	261
6.1.2	Variant prioritisation: variant class	262
6.1.3	Variant prioritisation: bioinformatics tools	268
6.1.4	Extreme phenotypes	272
6.2	Methods	273
6.2.1	Phenotypic data	273
6.2.2	Genetic data	275
6.2.3	Quality control and variant filtering.....	275
6.2.4	Covariates and population structure	279
6.2.5	Association analyses.....	279
6.2.6	Gene- and pathway-based analyses.....	280
6.3	Results.....	281
6.3.1	Quality control and variant filtering.....	281
6.3.2	Association analyses.....	282
6.3.2.1	Single variant approach	282
6.3.2.2	Gene and pathway-based approaches	286
6.4	Discussion.....	287
6.4.1	Power: study design and variant prioritisation.....	288
6.4.2	Results of interest	292
6.4.3	Rare variant detection with arrays versus sequencing	296
6.4.4	Rare variant/common disease hypothesis	298

6.4.5	Conclusions and future work	299
7	Exonic variation and response to the Ketogenic diet: whole exome sequencing.....	301
7.1	Introduction.....	301
7.1.1	Accurate variant detection	303
7.1.1.1	Avoiding errors generated in the library	304
7.1.1.2	Avoiding base-calling error	304
7.1.1.3	Avoiding alignment (or mapping) error	305
7.1.1.4	Calling variants	306
7.1.1.5	Variant filtering	310
7.1.2	Strategies for identifying susceptibility variants or genes in exome sequencing data	311
7.1.3	Pathway analysis	317
7.1.4	Population stratification	318
7.1.5	Interpretation and utility of results.....	320
7.2	Methods	321
7.2.1	Phenotypic data	321
7.2.2	Genotypic data.....	327
7.2.2.1	Quality control	329
7.2.2.2	Per-variant quality control.....	329
7.2.2.3	Per-individual quality control.....	330
7.2.3	Population structure	336
7.2.4	Association analyses.....	337
7.2.4.1	Single variant approach	337
7.2.4.2	Gene-based approach	337
7.2.4.3	Correction for multiple testing	341
7.2.4.4	Pathway-based analysis.....	341
7.3	Results.....	342
7.3.1	Quality control and population structure.....	342
7.3.2	Association studies.....	346
7.3.2.1	Single variant approach	346

7.3.2.2	Gene-based approach	348
7.3.2.3	Pathway analyses	358
7.4	Discussion.....	360
7.4.1	An overview of strategies adopted to increase power	361
7.4.2	Potential sources of error	364
7.4.3	Results of interest	368
7.4.3.1	Single variant and gene-based approach	368
7.4.3.2	Pathway-based approach	373
7.4.4	Conclusions and future work.....	378
8	Conclusions, study limitations and future work.....	380
8.1	Conclusions.....	380
8.1.1	Comments on phenotypic results	380
8.1.2	Comments on genotypic results	382
8.2	Study limitations.....	392
8.2.1	Classification of KD response.....	392
8.2.2	Compliance to treatment	393
8.2.3	Prospective versus retrospective recruitment	394
8.2.4	Sample size	395
8.2.5	Population structure	396
8.2.6	Replicates and batch effect	397
8.3	Future work.....	397
8.3.1	Continuing recruitment and further analyses	397
8.3.2	Gene expression analyses.....	399
8.3.3	Follow-up studies.....	400
8.3.4	Clinical translation.....	401
8.4	Final words	401
9	References	403
10	Appendices.....	461

Appendix 1.1	461
Appendix 1.2	462
Appendix 2.1	464
Appendix 2.2	467
Appendix 2.3	469
Appendix 2.4	487
Appendix 4.1	489
Appendix 4.2	490
Appendix 4.3	491
Appendix 5.1	492
Publications arising from this work.....	493

List of Tables

Table 1.1: Composition of Ketogenic diets and variants	29
Table 1.2: Studies detailing response to the Classical Ketogenic diet with 3-24 months follow-up	34
Table 1.3: Studies detailing response to the Medium Chain Triglyceride, Ketogenic diet, Modified Atkins diet and Low Glycaemic Index Treatment	38
Table 1.4: Conditions caused by genetic factors that have a 'probable' benefit from the Ketogenic diet.....	80
Table 1.5: Conditions caused by genetic factors that have a 'suggested' benefit from the Ketogenic diet.....	81
Table 3.1: Cohort clinical and demographic data (for cases with diet response data, n=253)	114
Table 3.2: Reported benefits on seizures, aside from reduced total seizure frequency, from the Ketogenic diet.....	123
Table 3.3: Reported benefits, independent of seizures, from the Ketogenic diet	124
Table 3.4: Results from Chi-square tests assessing independence of diet response at different time points	126

Table 3.5: Putative factors affecting 3-month Ketogenic diet response.....	129
Table 3.6: Putative factors affecting summary Ketogenic diet response.....	130
Table 3.7: Putative factors affecting extreme summary Ketogenic diet response.....	132
Table 3.8: Putative factors affecting diet response at 3-month follow-up: biochemical parameters	135
Table 3.9: Putative factors affecting diet response at 6-month follow-up: biochemical parameters	135
Table 3.10: Putative factors affecting diet response at 12-month follow-up: biochemical parameters	136
Table 4.1: Previously reported SNPs found in <i>KCNJ11</i> and <i>BAD</i> in Ketogenic diet cohort..	174
Table 4.2: R^2 values between pairs of previously reported <i>KCNJ11</i> and <i>BAD</i> SNPs found in Ketogenic diet cohort.....	177
Table 4.3: Observed and expected population genetics data for <i>KCNJ11</i> and <i>BAD</i> SNPs (from the 1000 Genomes Project) and results from Chi-square test of independence	178
Table 4.4: Breakdown of Ketogenic diet responders and non-responders by ethnicity, n=303	179
Table 4.5: Results of association analysis: common and intermediate variation in <i>KCNJ11</i> and <i>BAD</i> in Ketogenic diet responders and non-responders (n=303 for <i>KCNJ11</i> and n=246 for <i>BAD</i>)	181
Table 4.6: Results of association analysis: common and intermediate variation in <i>KCNJ11</i> and <i>BAD</i> in Ketogenic diet responders and non-responders, including ethnicity as a covariate (n=303 for <i>KCNJ11</i> and n=246 for <i>BAD</i>)	182
Table 4.7: KD response in cases with previously unreported variants or variants with $MAF < 0.01$ (in this cohort) in <i>KCNJ11</i> and <i>BAD</i>	184
Table 4.8: Top 10 most significant results for haplotype association analysis: variation in <i>KCNJ11</i> and <i>BAD</i> in Ketogenic diet responders and non-responders (n=303 for <i>KCNJ11</i> and n=248 for <i>BAD</i>)	187
Table 4.9: Results of association analysis, not accounting for ethnicity: common and intermediate variation in <i>KCNJ11</i> and <i>BAD</i> in Ketogenic diet overall responders and non-responders (summary KD response, n=216)	489
Table 4.10: Results of association analysis, including ethnicity as a covariate: common and intermediate variation in <i>KCNJ11</i> and <i>BAD</i> in Ketogenic diet overall responders and non-responders (summary KD response, n=216)	490

Table 4.11: Top 10 most significant results for haplotype association analysis: variation in <i>KCNJ11</i> and <i>BAD</i> in Ketogenic diet overall responders and non-responders (summary KD response, n=216)	491
Table 5.1: Genes with lowest p-values in 3-month and summary KD response gene-set analysis	238
Table 5.2: Known regions of long-range linkage disequilibrium (identified by Price et al [686]) excluded from Identity-by-State calculations	492
Table 6.1: Function of variants included in the Group 0 MAF<5% and MAF<0.5% analyses (prior to per-SNP quality control filtering)	281
Table 7.1: Description of Ketogenic diet response in 45 samples who had whole exome sequencing.....	323
Table 7.2: Mean performance metrics for responders and non-responders as measures of ancestry matching	337
Table 7.3: Function of variants included in the exome sequencing analyses	343
Table 7.4: Performance metrics for all exonic variants for all samples included in analyses, n=36	344
Table 7.5: Genes that reached significance or suggestive significance in c-alpha analyses: Groups 0 and 1.....	349
Table 7.6: Genes of most interest from exome sequencing analyses.....	355

List of Figures

Figure 1.1 ‘Metabolic pathways involved in Ketogenic diet treatment’, taken from Masino and Rho (2012)	49
Figure 1.2: ‘Hypothetical pathways leading to the anticonvulsant effects of the ketogenic diet’, taken from Bough and Rho (2007) [166]	52
Figure 1.3: The K_{ATP} -glycolysis hypothesis, taken from Ma et al., 2007	62
Figure 1.4: Genes shown to be upregulated by the Ketogenic diet in animal models	86
Figure 1.5: Genes shown to be downregulated by the Ketogenic diet in animal models.....	87
Figure 2.1: Phenotype proforma.....	100
Figure 3.1: Flowchart of cases recruited and included in final analysis.....	113
Figure 3.2: Distribution of diet compliance scores, n=66.....	117

Figure 3.3: Scatter plot of compliance scores and length of time on diet when questionnaire was completed, for 66 individuals	118
Figure 3.4: 3-month diet response, n=257 (n=252 with classifiable diet response data) ...	120
Figure 3.5: 6-month diet response, n=171 (n=160 with classifiable diet response data) ...	120
Figure 3.6: 12-month diet response, n=119 (n=111 with classifiable diet response data) .	121
Figure 3.7: 18-month diet response, n=77 (n=69 with classifiable diet response data)	121
Figure 3.8: 24-month diet response, n=49 (n=46 with classifiable diet response data)	121
Figure 3.9: Line graph showing trend of Ketogenic diet responder/non-responder status over time for five cases	125
Figure 3.10: Summary diet response, n=253 (n=221 with classifiable diet response data)	128
Figure 3.11: Extreme summary diet response, n=253 (n=108 with classifiable diet response data)	128
Figure 4.1: Flow-chart of cases with diet response data and included in final candidate gene analyses	166
Figure 4.2: Detectable relative risk and disease allele frequency curves for 3-month KD response cohort, with 80% power, assuming a disease prevalence of 0.5%, $\alpha=0.007$, 126 cases and control:case ratio of 1.4. Black line=recessive model; purple line=dominant model; orange line=co-dominant model.....	188
Figure 4.3: Power and sample size curves, using a co-dominant model, disease and marker allele frequencies of 0.017, assuming a disease prevalence of 0.5%, control:case ratio of 1.4 and α of 0.007. Black line=relative risk of 2.0; orange line=relative risk of 1.5; purple line=relative risk of 1.4.	189
Figure 5.1: Illustration of the concept of direct or indirect association in GWAS	199
Figure 5.2: Flowchart of cases recruited and included in final GWAS	212
Figure 5.3: Mean heterozygosity rates per individual, separated by ethnicity	214
Figure 5.4: For all individuals with Ketogenic diet response data: plot of values of the first and second vectors obtained by multidimensional scaling of pairwise identity-by-state values, with ethnicity highlighted	218
Figure 5.5: For all individuals with Ketogenic diet response data: plot of values of the first and second vectors obtained by multidimensional scaling of pairwise identity-by-state values, with 3-month response highlighted	219
Figure 5.6: Scree plot for the first 10 eigenvalues of the Multidimensional scaling matrix	220

Figure 5.7: Quantile-quantile plots of GWAS results from Fisher's exact test, using unadjusted p-values. A) 3-month diet response, $\lambda=1.00$; B) Summary diet response, $\lambda=1.00$ 228

Figure 5.8: Quantile-quantile plots of GWAS results from logistic regression, using unadjusted p-values. A) 3-month diet response, $\lambda=1.008$; B) Summary diet response, $\lambda=1.0002$ 229

Figure 5.9: Regional association plot and linkage disequilibrium (LD) map for rs12204701+/-500Kb. In the association plot, the left y-axis represents $-\log_{10}$ (P-values) for association with 3-month KD response; the right y-axis represents the recombination rate; the x-axis represents base-pair positions along the chromosome (human genome build 37). The top variant, rs12204701, is shown in purple, the rest of the variants are coloured according to their linkage disequilibrium r^2 value with rs12402021. In the LD map, LD is indicated as D'/LOD . The colour white = $LOD < 2$, $D' < 1$; blue = $LOD < 2$, $D' = 1$; shades of pink/red = $LOD \geq 2$, $D' < 1$; bright red = $LOD \geq 2$, $D' = 1$. LD pattern is based on genotype data obtained for this study. Confidence interval minima for strong LD: lower: 0.7, upper 0.98; Upper confidence interval maximum for strong recombination: 0.9; Fraction of strong LD in informative comparisons are at least 0.95; markers with <0.05 MAF are excluded 231

Figure 5.10: Regional association plot and linkage disequilibrium (LD) map for rs12402021+/-500Kb. In the association plot, the left y-axis represents $-\log_{10}$ (P-values) for association with 3-month KD response; the right y-axis represents the recombination rate; the x-axis represents base-pair positions along the chromosome (human genome build 37). The top variant, rs12402021, is shown in purple, the rest of the variants are coloured according to their linkage disequilibrium r^2 value with rs12402021. In the LD map, LD is indicated as D'/LOD . The colour white = $LOD < 2$, $D' < 1$; blue = $LOD < 2$, $D' = 1$; shades of pink/red = $LOD \geq 2$, $D' < 1$; bright red = $LOD \geq 2$, $D' = 1$. LD pattern is based on genotype data obtained for this study. Confidence interval minima for strong LD: lower: 0.7, upper 0.98; Upper confidence interval maximum for strong recombination: 0.9; Fraction of strong LD in informative comparisons are at least 0.95; markers with <0.05 MAF are excluded . 232

Figure 5.11: HapMap figure of rs12204701 +/-500Kb, including nearby Entrez genes and dbSNP SNPs 234

Figure 5.12: Detectable relative risk and disease allele frequency curves for 3-month KD response cohort, with 80% power, assuming r^2 of 0.9 between genotyped marker and causal variant, a disease prevalence of 0.5%, $\alpha=1.00 \times 10^{-7}$, 116 cases and control:case ratio of 1.01. Black line=recessive model, orange line=dominant model, purple line=co-dominant model 235

Figure 5.13: Power and sample size curves, using a recessive model (black line), dominant model (orange line) and co-dominant model (purple line), assuming $r^2=0.9$ between genotyped marker and causal variant, a disease prevalence of 0.5%, disease and marker allele frequencies of 0.2 and relative risk of 1.4, $\alpha=1.00 \times 10^{-7}$ and a control to case ratio of 1.4 236

Figure 5.14: A Snapshot of the Protein Interaction Database PRL Signalling events pathway, taken from http://pid.nci.nih.gov	245
Figure 6.1: Extreme responders and non-responders, n=106 (out of 250 individuals with diet response data and genotyped with the Illumina chip)	274
Figure 6.2: The Pathway Interaction Database S1P4 pathway, taken from http://pid.nci.nih.gov	287
Figure 7.1: Proportion of exonic bases achieving $\geq 2X$ coverage	331
Figure 7.2: Proportion of exonic bases achieving $\geq 10X$ coverage	332
Figure 7.3: Proportion of exonic bases achieving $\geq 20X$ coverage	332
Figure 7.4: Proportion of exonic bases achieving $\geq 30X$ coverage	333
Figure 7.5: Ti/Tv ratio of all exonic variants per individual	333
Figure 7.6: Mean genotype quality score per individual	334
Figure 7.7: Proportion of singletons per individual	334
Figure 7.8: Proportion of heterozygous genotypes per individual	335
Figure 7.9: Identity-by-state matrix: genotypic similarity between all pairs of individuals	336
Figure 7.10: BIOCARTA PTC1 Pathway, taken from http://www.biocarta.com/pathfiles ..	359
Figure 8.1: Diagram of the insulin/insulin-like growth factor 1/insulin receptor substrate signalling cascade, taken from Tao et al. 2007 [1140]	385

List of abbreviations

2DG	2-deoxy-D-glucose
ACA	acetoacetate
ADAS-cog	Alzheimer's Disease Assessment Scale-Cognitive subscale
ADP	adenosine diphosphate
AED	anti-epileptic drug
ATP	adenosine triphosphate
BCH	Birmingham Children's Hospital
BDNF	brain-derived neurotrophic factor
BHB	beta-hydroxybutyrate
BMI	body mass index
BRHSC	Bristol Royal Hospital for Sick Children
BWA	Burrows-Wheeler Aligner

CI	confidence interval
CNS	central nervous system
CNV	copy number variant
CSF	cerebrospinal fluid
DEND	developmental delay, epilepsy, and neonatal diabetes
DNA	deoxyribonucleic acid
ECH	Evelina Children's Hospital
EDTA	ethylenediaminetetraacetic acid
EEG	electroencephalography
ESP	Exome Sequencing Project
GABA	gamma-aminobutyric acid
GERP	Genomic Evolutionary Rate Profiling
GLUT1	glucose transporter 1
GO	Gene Ontology
GOSH	Great Ormond Street Hospital for Children
GWAS	genome-wide association study
HDL	high-density lipoprotein
HIE	Hypoxic ischemic encephalopathy
HWE	Hardy-Weinberg equilibrium
IBD	identity-by-descent
IBS	identity-by-state
ILAE	International League Against Epilepsy
ION	Institute of Neurology
IQ	intelligence quotient
IQR	interquartile range
KD	Ketogenic diet
LCT	long-chain triglyceride
LD	linkage disequilibrium
LDL	low-density lipoprotein
LGIT	low-glycaemic index treatment
LGS	Lennox-Gastaut syndrome
LOD	logarithm of the odds
LRT	likelihood ratio test
MAD	Modified Atkin's diet
MAF	minor allele frequency
MARS	Medication Adherence Report Scale
MCM	minichromosome maintenance
MCT	medium-chain triglyceride
MDS	multidimensional scaling
NCBI	National Center for Biotechnology Information
NHLBI	National Heart, Lung, and Blood Institute
NHNN	National Hospital for Neurology and Neurosurgery
NHS	National Health Service
NICE	National Institute for Health and Clinical Excellence

mRNA	messenger ribonucleic acid
PCA	principal component analysis
PCR	polymerase chain reaction
PDD	pyruvate dehydrogenase deficiency
PHHI	hyperinsulinaemic hypoglycaemia of infancy
PID	Pathway Interaction Database
PNDM	permanent neonatal diabetes
PUFA	polyunsaturated fatty acids
RCHM	The Royal Children's Hospital in Melbourne
RCT	randomised controlled trial
RNA	ribonucleic acid
ROS	reactive oxygen species
SIFT	Sorting Intolerant From Tolerant
SJS	Stevens-Johnson syndrome
SNP	single nucleotide polymorphism
SNV	single nucleotide variant
TENS	toxic epidermal necrolysis
UCL	University College London
UK	United Kingdom
UTR	untranslated region
VNS	vagus nerve stimulation
YE	Young Epilepsy

Acknowledgements

First and foremost, I would like to thank my supervisors, Professors Sanjay Sisodiya and Helen Cross, for giving me the opportunity to have this academic experience. It has been a steep learning curve, which I have truly enjoyed and appreciated.

To the clinicians and health care professionals at all study sites, especially those at Great Ormond Street Hospital, Evelina Children's Hospital and Matthew's Friends: thank you for allowing me to invade your clinics and for putting up with my incessant pestering. The Keto dietitians, more than anybody, have propelled my love for all things high-fat, low-carb.

I am extremely grateful to Dr Jon White and Dr Costin Leu for many insightful discussions and helpful advice regarding genetics and programming. Also to my friends and colleagues at the Institute of Neurology and Institute of Child Health for showing me the ropes and making me feel very welcome.

Finally, I would like to say a massive thank you to my husband, Andreas, for your patience and words of encouragement. I know you will be delighted to see the back of this thesis, not least because it will mean an exponential rise in the baked goods produced from our kitchen.

This work is dedicated to the memory of those who, very sadly, are no longer with us.

1 Background: Epilepsy, the Ketogenic diet and Genetics

1.1 Epilepsy

1.1.1 Definitions

Epilepsy is defined by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy as:

a disorder of the brain characterized by an enduring predisposition to generate epileptic seizures and by the neurobiologic, cognitive, psychological, and social consequences of this condition. The definition of epilepsy requires the occurrence of at least one epileptic seizure [1].

The brain dysfunction may result from a variety of causes, and thus epilepsy is not thought to be a 'singular disease entity' [1].

A seizure, the clinical manifestation of epilepsy, is defined as 'a transient occurrence of signs and/or symptoms due to abnormal excessive or synchronous neuronal activity in the brain' [1]. Based on mode of onset and termination, and clinical manifestation [1], seizures may be classified as *generalised* ('originating at some point within, and rapidly engaging, bilaterally distributed networks'), *focal* ('originating within networks limited to one hemisphere') or *unknown* [2]. Epileptic spasms are classified as *unknown*, as there is 'inadequate knowledge to make a firm decision regarding whether spasms should be classified as focal, generalised, or both' [2].

The epilepsies may be categorised according to their aetiology: *genetic*, where the epilepsy is 'the direct result of a known or presumed genetic defect(s) in which seizures are the core

symptom of the disorder', *structural/metabolic*, where there is a 'distinct other structural or metabolic condition or disease' associated with epilepsy, or *unknown* [2].

Epilepsy syndrome, or electroclinical syndrome, refers to 'a group of clinical entities that are reliably identified by a cluster of electroclinical characteristics' and other features, such as age of onset, EEG (electroencephalography) characteristics and seizure type; *constellations* refer to entities that are 'clinically distinctive' to electroclinical syndromes due to specific lesions or other causes; *epilepsies associated with structural or metabolic conditions* do not fit a 'specific electroclinical pattern' and may be organised by factors such as type of lesion, age of onset, localisation, seizure type, and specific ictal and interictal EEG patterns; other epilepsies are termed *of unknown cause* [2].

1.1.2 Burden

There are an estimated 70 million people worldwide who live with epilepsy [3, 4], of which approximately 10.5 million are aged under 15 years [5]. The median prevalence of active epilepsy (in which the last seizure occurred in the previous 12 months) is thought to be 7 per 1,000 for people of all ages [3]; the median prevalence for under-15s is estimated at 4.7 per 1,000. 20-40% of people with epilepsy worldwide are thought to have a drug-resistant condition [6], defined as 'failure of adequate trials of two tolerated and appropriately chosen and used AED [antiepileptic drug] schedules (whether as monotherapies or in combination) to achieve sustained seizure freedom' [7].

The total cost of epilepsy (including direct healthcare costs, direct non-medical costs and indirect costs) in Europe in 2010 was €13.8 billion [8]; a strong association has been identified between seizure frequency and cost in the USA [9] and Europe [10].

Financial burden is only one of the many costs of epilepsy. Adults with epilepsy have reported increased dissatisfaction with their education, family life, health, friends, social

life, achievement of their goals, and energy levels than adults without epilepsy [11].

Increased seizure frequency and severity have been associated with reduced health-related quality of life (QoL) [12]; seizure severity and seizure frequency may independently affect QoL [13]. Children with epilepsy have a relatively compromised QoL, in particular older children, those with a higher seizure frequency, those receiving polytherapy, those with partial seizures, and those who have been receiving treatment for longer [14-17]. In children, pharmacoresistance to seizures has been associated with lower full-scale intelligence quotient (IQ) [18] and has been identified as a significant risk factor for mortality [19].

1.1.3 Treatment and management

As outlined in the National Institute for Health and Clinical Excellence (NICE) guidelines [20], the 'mainstay' of epilepsy treatment is AEDs, which are usually recommended after a second epileptic seizure. Which AED(s) are used depends on the individual's seizure type, epilepsy syndrome, co-medication and co-morbidity, lifestyle, and the preferences of the person and their family and/or carers [21]. AEDs may cause adverse side effects, including dizziness, drowsiness, weight gain, visual field defects, nephrolithiasis, and movement and behavioural disorders [22, 23].

Epilepsy surgery may be considered in certain patients, when the potential benefits of the procedure outweigh the risks (including potential side effects, such as memory problems and visual field defects [24, 25]) and surpass chances of remission with further drug manipulation [26].

For the drug-resistant population, particularly (but not exclusively) for individuals not eligible for or who are opposed to epilepsy surgery, non-pharmacological treatments need consideration. Psychological interventions, such as relaxation therapy, cognitive behaviour

therapy and bio-feedback, the Ketogenic diet (KD) and vagus nerve stimulation (VNS) are alternative treatment options.

Although psychological interventions have not been shown to affect seizure frequency, they have improved QoL in some adults and may be suitable for use in children and young people with drug-resistant focal epilepsy [21].

VNS may be considered as an adjunctive therapy in children, young people and adults with drug-resistant epilepsy and who are not suitable for surgery [21]. VNS has been shown to reduce seizure frequency by $\geq 50\%$ in 26% participants aged <18 years [27]; a statistically significant improvement in seizure severity has also been reported, along with improved mood (independent of seizure frequency) [28]. Side effects previously reported with long-term VNS treatment include voice alteration, coughing, paresthesia and headaches [29].

Guidelines created in 'resource-rich' countries, such as those from NICE, may not be appropriate or feasible in 'resource-poor' countries [30], due to, for example, limited funding, infrastructure and ethical issues [31]. Epilepsy care guidelines, such as the World Health Organisation mental health Gap Action Programme [32], created for use in non-specialist health settings, may not underline (or even mention) the importance of non-pharmacological treatment options for people with epilepsy.

The continuing prevalence of epilepsy, despite the (theoretical) availability of over 20 AEDs, highlights the need for improved outcomes for people with this condition: management of seizures without a reduction in QoL. Although they may also cause adverse side effects, non-pharmacological interventions may help drug-resistant cases. This study focuses on the KD; its role in epilepsy treatment is outlined in the following section.

1.2 The Ketogenic diet and Epilepsy

Following preliminary reports concerning the effects of fasting on seizure cessation [33], Wilder [34] aimed to mimic the state of starvation and produce ketosis with a high-fat, low-carbohydrate diet. This led to the introduction of the so-called 'Classical Ketogenic diet', typically with a 4:1 ratio of fat (grams) to protein and carbohydrate (grams), for people with drug-resistant epilepsy.

The initial enthusiasm for the KD was ousted by the discovery of diphenylhydantoin in 1938 and the advent of new, easy-to-administer AEDs. Although widely used, concerns were quickly raised regarding adverse side effects of these drugs, such as gingival hyperplasia from diphenylhydantoin [35].

In the 1970s, in an attempt to make the diet more palatable, Huttenlocher introduced the Medium Chain Triglyceride (MCT) KD on the premise that MCTs are more ketogenic per calorie, and the diet therefore allows a greater bulk of protein and carbohydrate [36, 37]. The MCT KD originally derived 60% of its calories from MCT oil. A modified MCT KD, designed to decrease gastrointestinal side effects, derived 30% of its calories from MCT oil and 30% from long-chain fats [38]. Nowadays, the proportion of energy from MCT fats is dependent on individual requirements and tolerance.

The classical KD was established as the treatment of choice for a) pyruvate dehydrogenase deficiency (PDD), where deficiencies in a catalytic component of the mitochondrial enzyme PD complex (which catalyses the conversion of pyruvate to acetyl CoA and thus plays a crucial role in energy metabolism), are overcome by the provision of an alternative source of acetyl CoA by the KD [39, 40]; and b) glucose transporter 1 (GLUT1) deficiency syndrome [41], where impaired glucose transport across the blood-brain barrier is overcome by the

use of ketone bodies (which don't rely on the glucose transporter system to enter the brain) as fuel by the brain.

The KD experienced a reemergence when a programme was aired on NBC-TV's *Dateline* in October 1994, based on Charlie Abrahams and his response to the diet [42]. Charlie's father founded *The Charlie Foundation* to facilitate the availability of information regarding the diet for other parents. Other charities dedicated to the KD cause have subsequently been created, such as *Matthew's Friends* and *The Daisy Garland*.

Since the early 2000s, the KD repertoire has expanded to include more 'relaxed' variant forms of the diet, including the Modified Atkins diet (MAD) and the Low Glycaemic Index Treatment (LGIT), aiming to provide increased flexibility and palatability. The MAD is based on a ratio of approximately 1:1, although this is not necessary in all meals, and includes 10-30g of carbohydrate/day with no restriction of fluids, calories or protein. It allows users more flexibility and does not require the weighing of food portions or an initial hospital stay [43]. The LGIT includes a higher proportion of carbohydrates (approximately 40-60g/day) than the classical KD, with 60% of calories taken from fat, but only permits carbohydrates with a Glycaemic Index of <50 relative to glucose. In this project, the acronym 'KD' will be used to refer to all KD types, unless otherwise specified.

A summary of the composition of the KD and variant diets is given in Table 1.1. Fine-tuning by the dietitian is customary.

Table 1.1: Composition of Ketogenic diets and variants

Diet	Ketogenic ratio	% Carbohydrate	% Protein	% Fat (LCT)	% Fat (MCT)
Classical KD	4:1 (or 3:1, 2:1 and so on)	4	6	90	0
MCT KD	equivalent to 3:1	19	10	11	60
Modified MCT KD	equivalent to 3:1	19	10	41	30
MAD	equivalent to 0.9-1:1	~10	~25	65	0
LGIT	n/a	10	30	60	0

KD, Ketogenic diet; MCT, Medium Chain Triglyceride; MAD, Modified Atkins diets LGIT, Low Glycaemic Index Treatment; LCT, Long Chain Triglyceride

Nowadays, the classical KD, MCT KD and MAD are offered in approximately 15 NHS

(National Health Service) hospitals across the UK to children with drug-resistant epilepsy.

Two centres have a service dedicated to adults. Prior to commencing KD treatment,

individuals are screened (with biochemical testing of blood and/or urine) for disorders of

fatty acid metabolism and organic acidurias that may lead to deterioration when lipids are

used as the primary energy source or when dietary intake of protein or certain amino acids

is not restricted. For example, an individual with medium chain acyl-CoA dehydrogenase

deficiency would present with low serum and urine carnitines and elevated serum levels of

medium-chain-length fatty-acid-derived acylcarnitines, such as octanoyl carnitine

(<http://www.fodsupport.org/mcad.htm>). Many other factors also lead to caution being

taken when considering whether to commence KD, such as a family history of

hyperlipidaemia, nephrolithiasis, or severe gastro-oesophageal reflux.

Unless seizures worsen after starting the KD, it is customary to follow dietary treatment for

approximately three months before considering discontinuation [151]. Follow-up clinics

occur approximately every three months for the first year, although contact with a dietitian

is more frequent). If effective and tolerated, the diet is usually followed for two years

before discontinuation is considered, although the diet may be continued for much longer [44].

Modern clinical studies demonstrating the diet's effectiveness have fuelled the continuing interest in the KD among scientific and clinical communities.

1.2.1 The Ketogenic diet – clinical reports of effectiveness

In the first systematic review of KD response, it was estimated that $\geq 50\%$ seizure reduction is achieved in 56% children and seizure freedom is achieved in 16% [45]. In a subsequent systematic review, the estimated rate for achieving seizure freedom was 15.6% with 33% achieving $\geq 50\%$ seizure reduction [46]. In a meta-analysis, the pooled odds ratio of achieving $>50\%$ seizure reduction among patients who followed the KD until the end point of the study (this varied between studies) relative to those who had discontinued the diet before the end point was 2.25 [47]; 52% of individuals who remained on the diet for variable periods of time (up to 24 months) were reported to have achieved $\geq 90\%$ seizure control and 24% achieved complete seizure control.

The systematic reviews and meta-analysis consisted entirely of uncontrolled and predominantly retrospective studies, as highlighted by the authors. It was unclear whether studies had been screened independently by more than one researcher for inclusion in either systematic review, which may have introduced bias; in the meta-analysis, studies were rated for strength of evidence by one of three neurologists. Authors of both systematic reviews claimed to be investigating use of the KD in childhood epilepsy, but some of the studies included participants aged >18 years. Aside from the need for prospective, controlled trials, the authors also underlined the fact that primary outcomes were not uniform across all studies, long-term follow-up was limited and the clinical characteristics of dropouts were seldom described.

Since the aforementioned reviews and meta-analysis, a randomised controlled trial (RCT) assessing effectiveness of the KD for drug-resistant epilepsy has been conducted. Children aged 2-16 years with at least daily seizures or more than seven seizures per week, who had not responded to at least two AEDs were included in the study. 38% of participants treated with a classical KD or MCT KD achieved >50% seizure reduction at three month follow-up, compared with 6% of controls; 7% of treated participants experienced a >90% seizure reduction compared with no controls [153]. There was no significant difference in the number of children who achieved >50% or 90% reduction between the classical KD and MCT KD groups [98]. Responder rates are thought to be lower than those previously reported due to the intent-to-treat approach adopted in the RCT.

An RCT evaluating the effectiveness of the MAD in children has also been conducted [48]. 52% of 50 children who received the MAD achieved >50% seizure reduction at 3-month follow-up, compared to 11.5% of 52 controls; 30% of the treatment group achieved >90% seizure reduction, compared to 7.7% of controls.

In a blinded RCT, 20 children with Lennox-Gastaut syndrome were fasted and randomised to receive a KD plus a saccharin drink, or a KD plus a glucose drink (designed to act as a non-ketogenic diet) [49]. After six days, participants were fasted and then they received the KD with the alternative drink. There was no significant difference in improvement seen between the two groups, although a median of 1.5 fewer seizures per day were reported in the saccharin arm. The lack of statistically significant results is thought to be because ketosis was not completely eliminated in the group who received the glucose drink and baseline seizure frequency was measured following the fasting period, which may have reduced seizure frequency more effectively than the KD.

Other randomised trials have been conducted, but they compared the effects of different levels of carbohydrate intake [50], ketogenic ratios [51], or introduction of the KD with or without a fasting period [52], as opposed to randomising patients to KD treatment or no change in treatment.

A PubMed search was conducted to identify other studies (prospective or retrospective cohort studies, or uncontrolled studies) assessing the effectiveness of the KD, with 3-, 6-, 12- and/or 24-month follow-up data. The keywords 'ketogenic diet' were used and articles published in English, Spanish and French were included. Studies reporting on the KD for status epilepticus and metabolic disorders were excluded, as were abstracts.

The results for studies on the classical KD can be found in Table 1.2, and for the MCT KD, MAD and LGIT in Table 1.3. The percentage of participants classed as 'responders' ($\geq 50\%$ seizure reduction or seizure-freedom) in each study has been calculated with an intent-to-treat approach. Diet response rates are similar to those calculated in meta-analyses/systematic reviews [45-47], although seizure-freedom rates may be higher than calculated here, as not all studies provided this information. There are a higher proportion of responders in cohorts following the classical KD compared to other diet types; this may be reflective of the population demographics – younger patients are often more likely to follow the classical KD, due to the soft texture of the foods that tend to be used and the perceived ease of compliance in this age group, compared to, for example, adolescents. There are too few studies on the other diet types (in particular the MCT KD and LGIT) to permit comparison of response rates.

58 further studies were found in which participants had followed the diet for varying lengths of time and response was summarised independent of follow-up time. These were excluded, as they do not allow for inter-study comparison.

Responder rates in studies with fixed follow-up points of <3 months (range 12 days-10 weeks) were similar to 3-month response rates: 102/176(58.0%) achieved $\geq 50\%$ seizure reduction, of which 34(19.3%) became seizure-free [49, 53-62].

Table 1.2: Studies detailing response to the Classical Ketogenic diet with 3-24 months follow-up

Study	Design	Number recruited	Age range at diet onset (years)	Participants with ≥50% seizure reduction, n(% of original number recruited); of which seizure-free, n(%)¥			
				3 months	6 months	12 months	24months
[63]	Prosp	45	21-52	-	20(44%) with 'reduction in number of fits'; 2(4%)	-	-
[64]	Prosp	100	0.3-16	54(54%); NR	48(48%); NR	32(32%); NR	-
[65]	Prosp	150	1-16	85(57%); 4(3%)	72(48%); 5(3%)	64(43%); 11(7%)	-
[66]	Prosp	51	1-8	28(55%); 6(12%)	27(53%); 6(12%)	20(39%); 5(10%)	-
[67]	Prosp	41	1-18	13(32%); 6(15%)	10(24%); 3(7%)	3(7%); NR	-
[68]	Prosp	74	NR	24(32%); 14(19%)	39(53%); 19(26%)	52(70%); 23(31%)	-
[69] (also reported in [70])	Prosp	13	1-19	10(77%); 4(31%)	9(69%); 3(23%)	8(62%); 4(31%)	5(38%); 2(15%)
[71]	Prosp	35	0.2-13	17(49%); 5(14%)	15(43%); 4(11%)	10(29%); 3(9%)	-
[72]	Retro	143	0.3-29	59(41%); 21(15%)	60(42%); 24(17%)	54(38%); 23(16%)	-
[73]	Prosp	56	1-23	21(38%); 6(11%)	15(27%); 4(7%)	5(9%); 0(0%)	-
[74]	Retro	24	1-15	-	17(71%); 4(17%)	11(46%); 4(17%)	-
[75]	Retro	23	0.4-2	14(61%); 3(13%)	13(57%); 3(13%)	13(57%); 3(13%)	-

[76]	Retro	45	12-19	-	14(31%); NR	13(29%); NR	-
[77]	Retro	124	Mean=4.1±1.5 in one group; 5.3 ±1.6 in the other	94(76%); 43(35%)	-	-	-
[78]	Retro	53	0.5-15	25(47%) with >90% seizure reduction	29(55%) with >90% seizure reduction	-	-
[52]	Prosp	48	1-14	30(63%); 10(21%)	-	-	-
[79]	Retro	20	3-9	-	-	13(65%); 2(10%)	-
[80]	Prosp	12	0.6-6	-	-	10(83%); NR β	-
[81]	Retro	199	0.5-17	123(62%); 70(35%)	115(58%); 66(33%)	82(41%); 50(25%)	-
[82]	Retro	12	0.7-18	-	11(92%); 3(25%)	-	-
[83]	Retro	71	0.6-20	44(62%); NR	33(46%); NR	23(32%); NR	-
[84]	Retro	26	2-13	7(27%); NR	4(15%); NR	8(31%); NR	-
[85]	Retro	13	0.33-4.	11(85%); NR	8(62%); NR	6(46%); NR	-
[86]	Retro	43	0.5-3	30(70%); 15(35%)	23(53%); 17(40%)	-	-
[87]	Retro	57	1-26	-	-	31(54%); NR	-
[88]	Prosp	48	1-12	30(67%); NR	-	-	-
[89]	Prosp	25	1-18	16(64%); 2(8%)	15(60%); 3(12%)	12(48%); 2(8%)	-
[90]	Retro	70	1-12	-	44(63%); NR	38(54%); NR	29(41%); NR
[91] (also reported in [92])	Prosp	18	2-15	12(67%); 4(22%)	-	-	-
[93]	Retro	106	0.3-14.8	-	67(63%); NR	60(57%); NR	-
[94]	Prosp	76	0.3-16	60(79%); 33(43%)	-	-	-
[95]	Retro	25	0.5-15.8	-	13(52%); 0(0%)	-	-
[96]	Retro	47	Mean 47.2 ±33.7months	29(62%); 21(45%)	-	-	-

[97]	Retro	118	0.3–15	89(75%); NR	84(71%); NR	-	-
[98]*	Prosp	73 on classical diet	2-16	18(25%); 1(1%)	18(25%); 1(1%)	13(18%); 4(5%)	-
[99]	Prosp	30	1-15	16(70%); 3(13%)	-	-	-
[100]	Prosp	8	25-36	2(25%); NR	-	-	-
[101]*	Retro	17 on classical diet	0.11-12	11(65%); NR	7(41%); NR	4(24%); NR	-
[102]	Prosp	27	0.75-5	16(59%); 3(11%)	13(48%); 4(15%)	10(37%); 5(19%)	-
[103]	Retro	50	1-14	33(66%); 9(18%)	-	-	-
[104]	Prosp	38	0.3-5	27(71%); 11(29%)	28(74%); 12(32%)	20(53%); 9(24%)	-
[105]	Prosp	104	Mean 1.2±0.1	66(63%); 19(18%)	67(64%); 29(28%)	80(77%); 31(30%)	80(77%); 34(33%)
[106]	Prosp	12	24-65	5(42%); 2(17%) √	-	-	-
[107]	Prosp	48	1.3-12.2	26(54%); 10(21%)	-	-	-
[108]	Prosp	15	4-11	8(53%); NR	8(53%); NR	5(33%); 1(7%)	-
[109]	Retro	26	19.5±2.2 months	22(85%); 9(35%) ≠	21(81%); NR ≠	16(62%); NR ≠	-
[110]	Prosp	38	0.7-5	23(61%); 9(24%)	-	-	-
[111]	Prosp	26	1.3-15.8	17(65%); 5(19%)	-	-	-
[112]	Prosp	61	Median 4.8	29(48%); 2(3%)	24(39%); 2(3%)	19(31%); 2(3%)	-
[113]	Retro	71	1.5-18	36(51%); 3(4%)	20(28%); 1(1%)	17(24%); 1(1%)	-
[114]	Prosp	317	2 months to 17 years 8 months,	111(35%); 66(21%)	83(26%); 43(14%)	59(19%); 34(11%)	-
[115]	Prosp	17	mean age of 9.4±1.1 months	14(82%); 11(65%)	14(82%); 5(29%)	-	-
[116]	Prosp	31	0.6-7	22(71%); 14(45%)	-	-	-
TOTAL (not including [78] or [63])		3020	-	1402(46%); 441(15%)	1059(35%); 263(9%)	811(27%); 217(7%)	114(4%); 36(1%)

Prosp - prospective; Retro - retrospective

¥ number of seizure-free participants is included in the number with ≥50% seizure reduction

NR - not reported

* studies assessing effectiveness of more than one type of Ketogenic diet. Response rates given for participants on classical KD only

✓ Response rate at 4 months

≠ Response rates at 1-3months, 5-7months, 10-13months

β Response rates at 12 and 18 months compared to baseline

Table 1.3: Studies detailing response to the Medium Chain Triglyceride Ketogenic diet, Modified Atkins diet and Low Glycaemic Index Treatment with 3-24 months follow-up

Study	Design	Diet type	Number recruited	Age range	Participants with ≥50% seizure reduction, n(% of original number recruited); of which seizure-free, n(%)¥			
					3 months	6 months	12 months	24months
[117]	Prosp	MCT	15	1.5-9	9(60%); 1(7%)	6(40%); 2(13%)	3(20%); 1(77%)	-
[98]*	Prosp	MCT	72 on MCT	2-16	21(29%); 1(1%)	14(19%); 0(0%)	16(22%); 3(4%)	-
[118]	Prosp	MAD	20	3-16	14(70%); 3(15%)	12(60%); 3(15%)	-	-
[119]	Prosp	MAD	14	2-14	7(50%); 4(29%)	5(36%); 3(21%)	-	-
[50]	Prosp	MAD	20	3-16	7(35%); NR	10(50%); 2(10%)	-	-
[120]	Prosp	MAD	8	30-54	-	1(12.5%); 0(0%)	-	-
[121]	Prosp	MAD	30	18-53	14(47%); 1(3%)	7(23%); 1(3%)	-	-
[101]*	Retro	MAD	10 on MAD	0.4-15.2	2(20%); NR	2(20%); NR	1(10%); NR	-
[122]	Prosp	MAD	15	2-17	6(40%); NR	-	-	-
[123]	Prosp	MAD	5	4-18	3(60%); 1(20%)	3(60%); 1(20%)	-	-
[124]	Prosp	MAD	51	1-16	16(31%); 5(10%)	-	-	-
[125]	Prosp	MAD	38	1-15	17(45%); 5(13%)	13(34%); 0(0%)	9(24%); 0(0%)	-
[126]	Prosp	MAD	18	18-55	2(11%); 0(0%)	4(22%); 0(0%)	3(17%); 0(0%)	-
[127]	Retro	MAD	87	<18	-	36(41%); NR	28(32%); NR	8(9%); NR
[48]	Prosp	MAD	102 (50 in diet group)	2-14	26(52%); 5(10%)	-	-	-

[128] (includes data from [129])	Retro	LGIT	76	1.5-22	50% of 'the population'; NR	54% of 'the population'; NR	66% of 'the population'; NR	-
[130]	Retro	LGIT	15	1-20	4(27%); NR	7(47%); NR	6(40%); NR	4(27%); NR
[131]	Prosp	Classical and MCT	15	18-41	1(7%); 0(0%) √	1(7%); 0(0%)	2(13%); 0(0%)	-
[132]	Prosp	Classical, MCT and MAD	33	1-40	5(15%); NR	6(18%); NR	-	-
TOTAL (not including [128])			542		154(26%); 26(4%)	127(23%); 12(2%)	68(13%); 4(1%)	12(2%); NR

Prosp – prospective; Retro - retrospective

¥ number of seizure-free participants is included in the number with ≥50% seizure reduction

* studies assessing effectiveness of more than one type of Ketogenic diet. Response rates given for participants on specified diet type only

√ Response rate at 4 months

As for studies with fixed long-term follow-up points: 2/2 participants following a KD for 36 months achieved $\geq 50\%$ seizure reduction, one of whom was seizure-free [70]; of 15 people following the diet for >48 months, 13(86.7%) were responders, one of whom was seizure-free [133].

Evidence suggests that a favourable response to the KD may be sustained after discontinuation of dietary treatment. From the nine studies identified with follow-up after diet discontinuation in participants who had achieved seizure-freedom on the KD, 154/199(77.4%) remained seizure-free [96, 133-138]. Follow-up after discontinuation of the diet (not stated in three studies) ranged from 0-7 years.

Only one study looked into response following discontinuation of the KD in adolescents and adults: all three subjects who achieved $>75\%$ seizure reduction on the diet returned to pre-treatment frequency 1-7 months following cessation of treatment [106]. This may represent a disparity between dietary therapy in children and older people.

This review highlights the fact that dietary treatment, although effective in terms of seizure reduction in many patients with drug-resistant epilepsy, is not a panacea. Further trials of the MCT KD, MAD and LGIT with clearly defined follow-up times are needed. Studies with clearly-defined long-term follow-up times are also needed for all KD types.

It is unknown why the KD can have potential long-lasting effects on seizure control.

Additional studies are warranted to determine whether the antiepileptic effects of the diet outlast treatment, particularly in adolescents and adults.

1.2.2 Effects of the Ketogenic diet beyond seizure frequency

Other measures of treatment response that may lead to improved quality of life, independent of seizure reduction, must be considered.

The KD may lead to improvements in seizure severity in children and encourage continuation of dietary treatment despite inadequate seizure control [92, 118]. Seizure severity was not affected in one study involving adults following the KD [131], although anecdotal reports suggest otherwise (personal communication with metabolic dietitians from the National Hospital for Neurology and Neurosurgery, and Susan Wood, adult dietitian at Matthew's Friends Clinics for Ketogenic Dietary Therapies).

The KD has contributed to improved psychological function and quality of life in adults [100, 131, 139]; improvements were often independent of seizure control achieved on the diet and have also been reported in subjects without epilepsy [140-142]. Children following the KD may also have positive side effects, such as improved behaviour, sleep, alertness/attention, comprehension and motor skills [48, 73, 92, 143-148], or simply improved QoL [91, 149]. Improvements may be unrelated to seizure reduction [92, 143, 146, 148] or decreases in concomitant AED usage [146] and can result in a higher probability of continuing on the diet, more so than reduced seizures or medication [147].

The KD has been shown to slightly worsen psychosocial adjustment and mood (particularly anger/hostility and fatigue/inertia) in school-aged children and adolescents [148]. The authors attributed the decrease in psychosocial adjustment to the higher mean age of diet onset of patients in their cohort, compared to other studies. There are reports of dietary treatment interfering with social interaction in adolescents [76, 150].

The effects of the KD may reach further than solely seizure control, providing additional motivation for initiation or continuation of treatment. Such benefits are not consistently reported in studies investigating KD effectiveness, and further work is required to assess the proportion of patients that may profit from dietary treatment in this way. Equally, negative psychosocial effects from the diet should be reported in order to facilitate

prospective identification of patients at risk. Perhaps diet effectiveness should be defined by parameters other than seizure reduction alone.

1.2.3 Tolerability

Like most medical treatments, the KD may cause adverse side events. Such effects include gastrointestinal symptoms, hyperuricaemia, hypocalcaemia, hypomagnesaemia, acidosis, hypercholesterolaemia, renal stones and stunted growth [151].

In a meta-analysis of studies assessing the effectiveness of the KD in paediatric patients, the most frequently reported side effects were constipation (occurring in 14% of children remaining on the diet for at least three months); weight loss, growth problems, or anorexia (13%); nausea and vomiting (5%); behavioural problems or irritability (4%); increased serum cholesterol or triglycerides (4%); lethargy (4%); hypercalciuria (2.5%); increased liver enzymes (2.4%); renal stones (1.9%); diarrhoea (1.6%) and hypoglycaemia (1.3%) [152].

In the only RCT of the classical KD and MCT KD in a paediatric population, the most common side effect was constipation, reported by 33% of participants during the first three months of treatment [153]. One child had haematuria and renal echogenic debris whilst on the diet, which was successfully managed with potassium citrate. Other side effects included vomiting (24%), diarrhoea (13%), abdominal pain (9%), lack of energy (24%) and hunger (22%). There were no significant differences in tolerability between the classical KD and MCT KD groups [98].

In the RCT of the MAD, the most common side effect was constipation, reported by 46% of participants [48]. Other side effects included anorexia (18%), lethargy (6%) and vomiting (10%). Two children developed frequent chest infections, and another developed hyperammonaemic encephalopathy one week after starting the diet.

No direct comparison between adolescents or adults and children has been made of diet tolerability. From the limited data, side effects appear similar in all age groups [154].

Long-term complications of the KD are relatively unknown. A higher risk of fractures, renal stones and stunted growth has been found in children who followed the diet for >6 years [155]. Evidence appears most consistent for a negative effect on height and/or weight during KD implementation, particularly when the diet is started at a young age [137, 156-160]; some studies, however, have reported little or no effect on growth parameters [143, 158, 161]. Patients who follow the KD long-term may experience catch-up growth (height and weight) after discontinuation of dietary treatment, with most improvement in weight observed in ambulatory individuals, and most improvement in height and weight in individuals who achieved seizure-freedom on the KD. Abnormal lipid values may also gradually return to normal over time (2+ years) [44, 137, 162].

It has been recommended that signs of hyperlipidaemia, cardiac disease, linear growth failure, gastrointestinal disorders, nephrolithiasis, electrolyte/vitamin/ mineral/trace element deficiencies, and osteoporosis/osteopenia/fractures are monitored closely in patients following the KD long-term [163].

Adverse effects from short-term use of the KD are common, but mostly non-critical and/or transient. The risks associated with long-term use, however, are largely unknown.

1.2.4 Retention

46 studies from Table 1.2 and Table 1.3 gave initial study recruitment figures and retention rates at fixed follow-up points (including prospective and retrospective studies). 79% of the total number of participants who started the classical KD were still following the diet at the 3-month point; for studies that gave 6-month point retention figures, 63% remained on the diet, 43% remained at the 12-month point and 43% at the 24-month point (only two

studies). Retention rates for the MAD were 76% at the 3-month point, 64% at 6 months, 44% at 12 months and 18% at 24 months (one study). Only two studies provided retention information for the MCT diet: 70% continued on the diet at the 3-month point, 53% at 6 months, and 39% at 12 months. Only one study provided retention information for the LGIT: 67% continued on the diet at the 3-month point, 60% at 6 months, 47% at 12 months, and 40% at 24 months.

In adolescents and adults, retention rates may be even lower, although reports are inconsistent and sample sizes limited. Up to 78% of participants have discontinued dietary treatment even before three months [100]. Many may refuse to start the diet in the first place; in one study, only 18 out of 130 eligible adults consented to begin the KD [126].

Reasons for discontinuation of the KD are not consistently reported. In the RCT of the classical KD and MCT KD, of the 10 children who discontinued dietary treatment before the 3-month point, 9(90%) of these withdrew due to lack of tolerability: three because of parental unhappiness with the restrictions, two with behavioural food refusal, one with extreme drowsiness, and three due to gastrointestinal side effects; the remaining child withdrew due to increased seizures [98, 153]. In the RCT of the MAD, treatment was discontinued in the two children who developed frequent chest infections and one who developed hyperammonaemic encephalopathy [48]. Only one other child discontinued treatment prior to the 3-month point; this was due to the perceived restrictiveness of the diet by the child and family.

In a study exploring the reasons for discontinuation of dietary treatment prior to the 6-month point, 10/19(52.7%) children withdrew from the diet for medical reasons (lack of efficacy, complications or unrelated hospitalisation), and 9/19(47.4%) withdrew for non-medical reasons (caregiver issues or patient issues, such as food refusal) [150].

Discontinuation for medical reasons was more common in children <6 years of age, whereas non-compliance was more common in older children.

Other reports outline similar reasons for diet discontinuation: in one study with long-term follow-up, 33/67(49.3%) families discontinued the diet before 12 months due to ineffectiveness, 12/67(17.9%) due to intercurrent illness unrelated to the diet, and 20/67(29.9%) found it to be too restrictive [133]; in another study, of 183 patients who discontinued the diet after 0.03–12.3 years, 23% discontinued due to seizure freedom, 23% due to prolonged KD use with limited continued benefit, 20% due to diet ineffectiveness, 12% due to reduced efficacy over time, 11% due to medical illness, 8% due to perceived restrictiveness, and 3% due to worsening seizures [164].

Levels of compliance must also be considered alongside retention rates, as it may not always be easy or even possible for a patient to consistently adhere to a restrictive regime, depending on food preference, social circumstance and caregiver understanding of the KD. There is a tendency for studies to simply provide numbers/proportions of patients who are (allegedly) following the diet and of those who have discontinued. Levels of compliance have been noted in two studies, measured by communication with parents/patients [73] or using serum beta-hydroxybutyrate (BHB) or urine acetoacetate (ACA) [106]. Inter-individual variability in ketone body production and maintenance may be high [165], and response to the diet may not solely depend on ketosis *per se* [166]. Continual recording of food intake, as previously undertaken in a case report [167], may provide more accurate levels of compliance with the diet but would not be feasible on a wider scale or in outpatient settings. Alternative ways of assessing compliance are needed.

KD retention levels seem poor, although studies with long-term KD retention data are limited. The most common reason for discontinuation of treatment may be lack of

effectiveness, but the restrictiveness of the regime also affects dropout rates. This highlights the need for an improved selection process for the KD and for further study into the mechanisms underlying its treatment effects.

1.2.5 Biochemical response to the Ketogenic diet and putative mechanisms of action

The biochemical changes induced in response to the KD are similar to those induced by starvation: principally, a change in metabolism from 'gluco-centric' (use of glucose as an energy substrate) to 'adipo-centric' (where ketone bodies and fatty acids act as energy substrates) metabolism [168]. The principal difference is that lean body mass and blood glucose levels are maintained with the KD (although blood glucose levels may drop below normal, they are then maintained at this level and should not continue to fall), whereas they are diminished with prolonged starvation.

During the early stages of starvation, or the post-absorptive state, in response to lower blood glucose levels, insulin concentrations decrease and glucagon increases. This has three main consequences: hepatic glycogenolysis and gluconeogenesis are stimulated and fatty acid synthesis from acetyl CoA and hepatic glycolysis are inhibited. All these effects of glucagon are mediated through activation of the cyclic AMP cascade.

Triacylglycerol from adipose tissue is broken down to give fatty acids and glycerol. Fatty acids are used by the liver and muscle as an energy source and glycerol is converted to glucose.

The glucose formed from glycogenolysis is transported for use in muscle and adipose tissue, although hepatic glycogen stores are depleted after approximately a day of fasting. Uptake of glucose is reduced due to low insulin levels and glucose oxidation is inhibited in muscle and adipose tissue, which causes increased release of lactate, pyruvate, and alanine for gluconeogenesis. As previously mentioned, fatty acids become the main energy

substrate for muscle and liver. These processes (glycogenolysis and the release of glucose by the liver, the mobilisation of fatty acids from adipose tissue, and the shift in the fuel used by muscle and the liver from glucose to fatty acids) help to stop blood glucose levels from falling any further.

The body also reduces its need for glucose. When fatty acids are oxidised, the acetyl CoA formed during this process is converted to HMG-CoA (via HMG-CoA Synthase, a rate-limiting enzyme that is only found in the liver) and then into ACA (which breaks down to acetone) and BHB. This is thought to occur several days after starvation begins. Ketogenesis is also thought to occur in the kidneys [169]. Normally, the acetyl CoA is oxidized to CO₂ and H₂O in the citric acid cycle but, due to the large quantities generated from the degradation of fatty acids, not all can be oxidised in the citric acid cycle, partially because the process of gluconeogenesis depletes the supply of oxaloacetate, which is essential for the entry of acetyl CoA into the citric acid cycle. Instead, the ketone bodies produced are transported to tissues that are able to use them as energy substrates, such as the brain. Most neurons do not contain enzymes to be able to oxidise fatty acids [170].

A certain level of glucose is still needed (red blood cells, for example, can only use glucose for energy as they lack mitochondria containing the enzymes of fatty acid/ketone body oxidation) and, although some glucose is provided by hepatic and renal gluconeogenesis, the process of converting glycerol to glucose is limited. If ketone bodies were not produced and used as fuels by some tissues, lean body mass would be broken down to obtain amino acids for conversion to glucose [171].

As starvation is prolonged, the brain begins utilising more ketones and less glucose until ACA and BHB become the primary fuels of the brain. After 3.5 days of starvation, the amount of glucose consumed by the human brain is approximately 75% less than normal

[172]. After 5-6 weeks of starvation, approximately 60% of the brain's energy supply is from ketone bodies [173]. At the same time, muscles use progressively more fatty acids as energy substrates, and hepatic gluconeogenesis and hepatic urea production (produced from the nitrogen released from amino acid breakdown for excretion) decrease. From clinical observation, this process of keto-adaptation is thought to take 2-4 weeks [168, 174]. The performance of trained endurance athletes was not found to differ after four weeks of following a low-carbohydrate diet [175], whereas earlier studies showed performance to be limited in the early periods of adaptation to such diets. Although studies of long-term consumption of low-carbohydrate (with or without additional fat intake) have been studied in healthy individuals [176], those who are overweight [177, 178] and those with Type II diabetes [179-181], no studies were found that systematically documented the process of keto-adaptation over a long period of time.

The KD aims to replicate this state of starvation, encouraging the production of ketone bodies with high dietary fat intake. The speed at which keto-adaptation occurs depends on the rate at which the diet is introduced, the ketogenic ratios used, and whether it is preceded by a fasting period.

The hallmark features of the KD, focusing on the effects of energy substrate utilisation in the brain, are outlined in Figure 1.1.

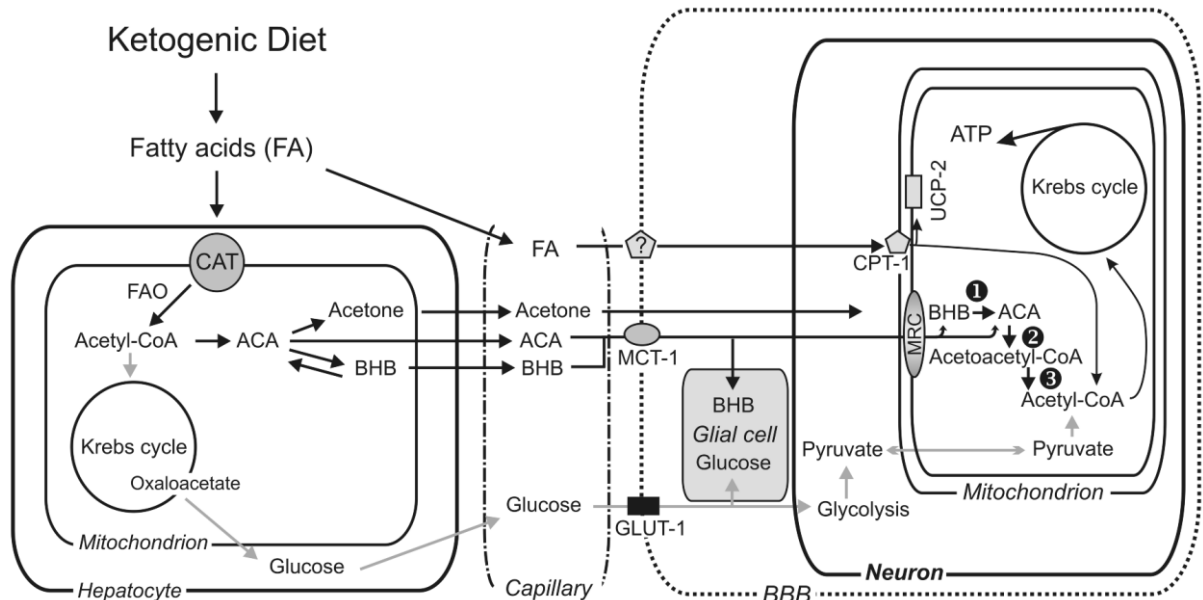


Figure 1.1 'Metabolic pathways involved in Ketogenic diet treatment', taken from Masino and Rho (2012)

This figure serves to show the complexity of the physiological changes that occur in response to the KD; this is only a snapshot, which does not include potential downstream effects of ketones or fatty acids in the brain. This network could be expanded to the entire body: without lipolysis from adipose tissue, the liver and kidneys would have insufficient energy to produce ketones or for amino acid oxidation for use for glucose synthesis, and if muscles did not preferentially use fatty acids as an energy source instead of glucose, less glucose would be available for tissues that are (partially or completely) dependent on it as an energy substrate. Theoretically, genetic variants affecting any one of these processes may influence KD response. The need to treat individuals with damaging mutations in *SLC2A1* (the gene encoding the glucose transporter protein type 1, which transports glucose across the blood-brain barrier) with the KD is well-documented, but there are many other potential candidate genes. It makes sense that the ability to transport ketone bodies across the blood brain barrier would be a predominant factor in determining the rate at which the brain can use ketone bodies. Thus, variants affecting the amount or function of the monocarboxylate transporters, MCT1 and MCT2, which are responsible for the transfer of ketone bodies across the blood-brain barrier and for their access to astroglial and

neuronal cells [182], would be of interest. However, considering that the capacity of ketone body transporter proteins is also dependent on plasma ketone concentration, variants affecting any link in the network that brings about an increase in ketone body production could also be relevant. Variants that influence the breakdown of BHB and ACA to acetyl CoA for entry into the tricarboxylic acid cycle (via D- β -hydroxybutyrate dehydrogenase and succinyl-CoA:acetoacetate-CoA transferase), or that influence levels of free fatty acids or blood glucose/glycolytic flux could also influence KD response, independent of ketone body levels. It is unknown exactly which of the physiological responses induced by the KD are most relevant for seizure control. If this was known, it would be easier to pinpoint potentially relevant genes.

Despite decades of endeavours to elucidate the underlying mechanisms of the KD, no definitive theory as to what causes its antiepileptic effects has been accepted. Having been designed to mimic the effects of starvation on the body, it may be assumed that the mechanisms underlying the antiepileptic/neuroprotective effects of calorie restriction/starvation are also responsible for those underlying the KD. Many postulated mechanisms of action do overlap, such as reduction of reactive oxygen species (ROS), antiinflammatory effects, increased mitochondrial biogenesis and neurotrophic factor activity [183], AMP kinase activation and subsequent inhibition of mammalian target of rapamycin (mTOR) [184], and reduced blood glucose and energy made from glycolysis [185]. Indeed, some studies suggest that calorie restriction alone is enough to have an antiepileptic effect [186-188] and that the effectiveness of the KD is improved when administered with fasting or calorie restriction [189, 190].

However, evidence suggests that there may be different underlying mechanisms of action. Only calorie-restricted KD-fed rats (as opposed to normal calorie-restricted or *ad libitum* diets) were protected against prolongation of 'kindling-like afterdischarges' in a maximal

dentate activation test [187]. Calorie restriction and the KD have had different effects on seizure thresholds in various seizure tests: the KD protected from 6 Hz-induced seizures and failed to protect against kainic acid-induced seizures, whilst the calorie-restricted diet caused increased seizure activity in the 6 Hz test, but protected juvenile mice against kainic acid-induced seizures [191]. KD-fed mice displayed a pattern of hepatic gene expression distinct from those fed a calorie-restricted diet [192].

Putative theories regarding how the KD exerts its antiepileptic effects, culminating in a network of complex interrelationships, are illustrated in Figure 1.2.

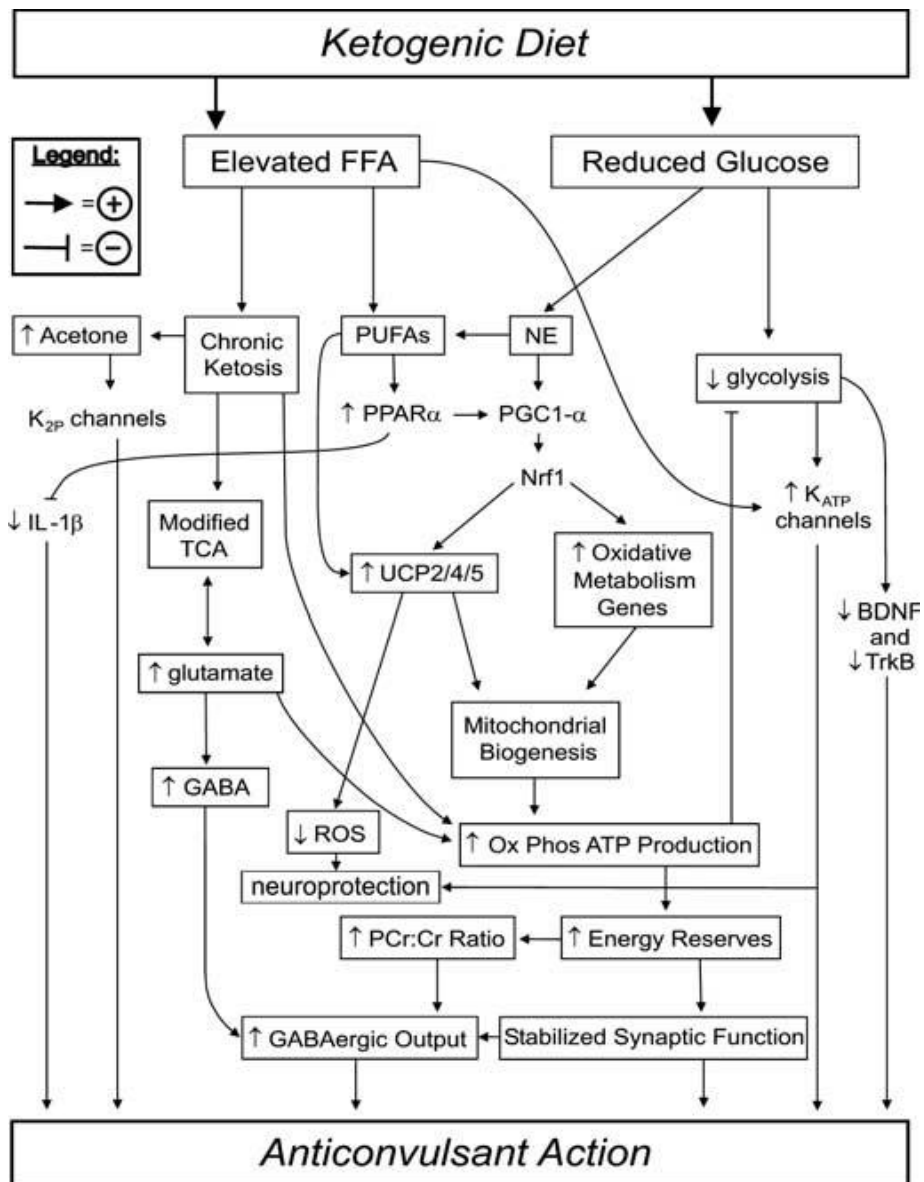


Figure 1.2: ‘Hypothetical pathways leading to the anticonvulsant effects of the ketogenic diet’, taken from Bough and Rho (2007) [166]

Postulated mechanisms of action of the KD are outlined below:

i) Ketone bodies

In an individual following the KD, a shift occurs in hepatocytes: the high level of acetyl CoA generated from fat metabolism is converted to the ketone body ACA and its derivatives, BHB and acetone. These ketone bodies serve as a cellular energy source, and can cross the blood-brain barrier.

Given elevated serum ketones levels in patients [193] and animals [194] following the KD, and the rapid reversal of seizure protection with glucose infusion [36], it is not surprising that much work has concentrated on the role of ketosis in the therapeutic mechanisms of the KD.

ACA [195, 196] and acetone [196-198] have been shown to have direct antiepileptic effects in animal models. *In vitro* studies have also revealed potential antiepileptic properties of ketone bodies: BHB and ACA slowed the firing rate of substantia nigra pars reticulata neurons [199]; ACA inhibited vesicular glutamate release in cultured neurons and sliced brain tissues, but not in astrocytes, reduced the amplitude of miniature excitatory postsynaptic currents in hippocampal slices, and suppressed the 4-aminopyridine-evoked seizures and glutamate release [200]; a combination of ACA and BHB prevented glutamate excitotoxicity in both acutely dissociated neocortical neurons and in isolated neocortical mitochondria [201], and protected against glutamate toxicity in hippocampal cell lines and primary hippocampal neurons [202]; presence of BHB increased the open probability of single K_{ATP} channels in mouse dentate gyrus granule cells [203], the implications of which will be discussed in the *Energy metabolism* section.

Using proton magnetic resonance spectroscopy, cerebral acetone has been detected in 5/7 individuals with either 'partial' or 'full' seizure control on the KD [204].

Evidence is, however, conflicting: BHB and ACA did not alter synaptic transmission in rat hippocampal-entorhinal cortex slices and cultured hippocampal neurons [205], nor did they modulate gamma-aminobutyric acid-a ($GABA_A$) or ionotropic glutamate receptors in cultured neocortical neurons [206]. Chronic treatment with BHB failed to suppress evoked and spontaneous epileptiform activity in hippocampal slice cultures, despite protecting the

cultures against chronic hypoglycaemia, oxygen-glucose deprivation, and NMDA-induced excitotoxicity [207].

Furthermore, blood BHB or urinary ACA levels are not always associated with improved seizure control in humans [50, 51, 55, 80, 93, 100, 109, 119, 121, 122, 124, 126, 128, 132, 208-213] or animals [186, 194, 214-219] following a KD.

Although inconsistent, *in vivo* and *in vitro* evidence suggests that ketones have antiepileptic properties, but they may not directly mediate these effects [166]. Ketones may also be neuroprotective (for example, protection against glutamate excitotoxicity), but this does not necessarily translate into increased seizure protection.

ii) *Reduced glucose usage and reduced glycolysis*

The KD, by its very nature, makes use of energy substrates (ketone bodies) that are directly metabolised by mitochondria and thus bypass, and possibly even inhibit [199, 220] glycolysis.

The proposal that reduced glucose and glycolysis is another mechanism by which the KD exerts its antiepileptic properties is supported by several facts: glucose uptake, glucose metabolism and glycolysis increase during seizures [221, 222], glycolytic energy is needed to maintain synaptic activity [223], and the effects of the KD can be rapidly reversed upon infusion of glucose [224].

Cultured hippocampal neurons and astrocytes from *Bad*-knockout mice, which have reduced cellular glucose metabolism [225], showed reduced mitochondrial utilisation of glucose and elevated utilisation of BHB, consistent with KD-induced changes in metabolism [226]. *Bad*^{-/-} mice displayed resistance to behavioural and electrographic seizures and seizure severity was diminished. *Bad*^{S155A} (non-phosphorylatable) mice, which have

opposite effects on BAD's apoptotic activity, displayed the same phenotype, highlighting the role of glucose metabolism.

Other experimental studies have yielded conflicting results. Mice fed a KD or a calorie-restricted diet were found to have lower blood glucose levels than those fed a control diet, but the two dietary interventions had different effects on seizure thresholds in 6 Hz- and kainic acid-induced seizure tests [191]. These findings that blood glucose levels are not correlated with KD response have been supported [216, 217, 219, 227] and contradicted [185, 228] in other KD rodent models.

Clinical reports are similarly inconclusive: a correlation has been reported between lower blood glucose levels and increased KD effectiveness [128, 229], but others have found that blood glucose levels were not correlated with diet response [109, 230]. The association may be clouded due to varying accuracy and comparability between different methods of blood glucose measurement [231], and high variability in blood glucose levels, even in normal individuals [165].

Despite these inconsistencies, experimental studies have continued to implicate low glucose levels and reduced glycolysis in the antiepileptic effects of the KD.

Glucose restriction in neurons from rat hippocampal slices, with sufficient or high levels of adenosine triphosphate (ATP), resulted in significant hyperpolarisation and decreased neuronal excitability [232]. The authors reported that this was due to increased extracellular adenosine acting at adenosine A1 receptors (A1R) and opening K_{ATP} channels.

2-deoxy-D-glucose (2DG), a glucose analogue that transiently inhibits glycolysis by blocking phosphoglucose isomerase [233], has reduced seizures induced by 6 Hz [234, 235], audiogenic stimulation [235] and pilocarpine [236] in rodents. 2DG has also reduced interictal epileptiform bursts in rat hippocampal slices [235] and increased the post-

discharge threshold in kindled rats [237]. It has not, however, protected against maximal electroshock [235, 236], pentylenetetrazole-induced [235, 236], kainic acid-induced [236], or amygdala-kindled seizures [234]. In one study, 2DG decreased seizure threshold in pentylenetetrazole, kainic acid and electroshock seizure threshold tests [234].

2DG has been shown to increase [238] and decrease [237] hippocampal expression of genes encoding brain-derived neurotrophic factor (BDNF) and to decrease expression of the gene encoding its receptor, TrkB tyrosine kinase [237] in rodents. 2DG also prevents or reduces the seizure-induced upregulation in expression of *Bdnf* and *Ntrk2* [237].

Diverting glucose metabolism from glycolysis to the pentose phosphate pathway with fructose-1,6-bisphosphate provides protection against seizures induced by pilocarpine, kainic acid or pentylenetetrazole in rats [236].

iii) Fatty Acids

The antiepileptic properties of fatty acids, in particular polyunsaturated fatty acids (PUFAs), have made them prime candidates in the search for underlying mechanisms of action of the KD.

Animal studies have demonstrated that PUFAs have antiepileptic activity [239-247], although others have not found this to be true for eicosapentaenoic and docosahexaenoic acid [248], or linoleic acid and alpha-linolenic acid [249]. Clinical studies have supported the idea that docosahexaenoic acid has antiepileptic properties [250, 251], but findings are conflicting [252, 253].

Supplementation of a standard diet with linoleic and alpha-linolenic acid can exert anticonvulsive properties comparable to a KD [243].

One potential way in which PUFAs may exert an antiepileptic effect is through peroxisome proliferator-activated receptor (PPAR) activation. These nuclear receptor hormones are transcription factors that bind to DNA and regulate gene expression, and also serve as intracellular receptors by binding lipids [254]. Dietary PUFAs, in particular alpha-linolenic acid and linoleic acid, are thought to be potent PPAR α ligands [255]. PPAR α -activated promotion of transcription of genes, such as that encoding mitochondrial 3-hydroxy-3-methylglutaryl-CoA synthase (*Hmgcs2*) in the brain and liver, is thought to be at least partly responsible for the rise in blood ketone levels with the KD [256, 257].

Hepatic PPAR α is also involved in the regulation of a variety of genes encoding enzymes of intermediary metabolism and neurotransmitters, such as the glutamic-oxaloacetate transaminases, cyclooxygenase 2 and inducible nitric oxide synthase [256]. This adaptation mimics the transcriptional response to fasting [258] and mirrors one proposed mechanism of some AEDs, such as valproate: the alteration of neurotransmitter concentrations implicated in epilepsy [259].

Fatty-acid-stimulated PPAR α , PPAR δ and PPAR γ (together with the co-activator, peroxisome proliferator-activated receptor γ co-activator 1) have been implicated in the control of gene expression of uncoupling proteins (UCPs) [260, 261]. These are mitochondrial transporters present in the inner mitochondrial membrane that dissipate the proton gradient generated by the respiratory chain so that heat is generated, instead of ATP [262, 263]. *In vitro* experiments show that increased expression of *Ucp2* protects against neuronal cell death in the face of toxicity [264-267]. Levels and activity of *Ucp2*, *Ucp4* and *Ucp5* were increased in the hippocampi of mice fed a KD, compared to those fed a standard diet [268]; this was also associated with decreased ROS production. Decreased ROS is just one of the numerous consequences of increased UCPs, which may be (in part) responsible for the neuroprotective/antiepileptic effects of the KD; others include

increased mitochondrial biogenesis, increased ATP levels, decreased lipid peroxidation, regulation of mitochondrial calcium flux and heat generation in presynaptic nerve terminals [264, 269].

Other proposed mechanisms for the modification of seizure threshold by PUFAs include the inhibition of voltage-gated sodium and calcium channels, activation of specific K_{2P} channels [166], a combined effect on A-type potassium channels, sodium channels and resting membrane potential [270], and antiinflammatory actions [271].

Clinical evidence regarding the significance of PUFAs in the KD is conflicting. Serum PUFAs were significantly increased in children following a KD, and a positive correlation between reduction in seizure frequency and elevation in serum total arachidonic acid was reported [212]. The change in total plasma fatty acids during KD treatment has been shown to mirror the effectiveness of the diet [272]. Others have found that changes in fatty acids levels in plasma phospholipids did not correlate with likelihood of >90% seizure reduction at the 6-month point, compared to prior to starting the diet [89]. Responders to the KD have been found to have a significant increase of serum palmitoleic acid and a significant decrease of arachidonic acid compared to non-responders, although no difference was found in levels of linoleic acid or alpha-linolenic acid [101].

The medium chain saturated fatty acid, decanoic acid, has been implicated in the mechanism of the MCT KD: it improved seizure control in an *in vitro* pentylenetetrazol model compared to valproate [273] and it has been shown to increase mitochondrial proliferation in a neuronal cell line [274]. Decanoic acid is elevated in patients during dietary treatment [275, 276], although levels were not found to differ between responders and non-responders.

Despite these theories regarding different types of fatty acids, KDs containing either MCT, flaxseed oil (rich in alpha-linolenic acid), butter, or an equal combination of the three, conferred a similar amount of seizure protection in rats, independent of differences in ketosis [215, 277]. Response rates in people following a classical KD or an MCT KD were found not to differ significantly [98, 278, 279]. Diet responders have been found to have larger absolute decreases in plasma phospholipid fatty acid 18:0 (number of carbons: number of double bonds) and lesser increases in 24:1 during dietary treatment, compared to non-responders [89]. No difference was found in levels of the fatty acids 14:0, 16:0, 16:1, 18:1, 18:2, 18:3, 20:0, 20:3, 20:4, 20:5, 22:0, 22:5, 22:6 and 24:0, measured pre-diet and during dietary treatment, between responders and non-responders.

There is little evidence to suggest that one particular type of fatty acid is required for the antiepileptic effects of the KD. The importance of the quantity of fatty acids is unclear.

iv) Reduced oxidative stress

In addition to the effects of UCPs on ROS (discussed in *Fatty Acids* section), there is evidence that ketone bodies reduce ROS generation [201, 268, 280-283]. This may be important, as ROS have been implicated in seizure generation and epileptogenesis [284, 285], and oxidative stress generated by ROS is associated with neuronal loss in neurodegenerative diseases [286].

Other measures taken from KD-fed animals are indicative of enhanced mitochondrial antioxidant status: elevated mitochondrial glutathione (increased activity and up-regulated protein levels of glutamate cysteine ligase indicated *de novo* glutathione synthesis), reduced hippocampal (but not in the frontal cortex) CoA and lipoic acid [282], and increased hippocampal (but not cerebellar) glutathione peroxidase activity [287], compared to controls.

Tumour tissue from animals fed a KD displayed a differential expression of specific genes involved in the regulation of ROS levels, to mirror that seen in non-tumour brain tissue [283]. Although this study was not concerned with the antiepileptic effects of the KD, neuroprotective changes seen in tumours, such as reduced Cox2, increased glutathione peroxidase 7, and increased serine peptidase inhibitor, may also be relevant for protection against seizures.

Treatment of human embryonic kidney cells with BHB was found to inhibit class I histone deacetylases and, in mouse kidney, it lead to increased transcription of oxidative stress resistance genes, such as those in the FOXO3A network (*Foxo3a*, *Mt2*, *Lcn2*, *Lemd3* and *Hbp1*), Mn-SOD, and catalase [270].

v) *Energy metabolism*

Neuronal energy consumption is substantial, predominantly in order to maintain resting potentials and for neuronal signalling [288]. One might expect more efficient/increased energy metabolism to lead to stabilised synaptic function, limited neuronal hyperexcitability and fewer seizures.

It has been shown that rats fed a high-fat diet for three weeks had increased 'cerebral energy reserve', with an increased cerebral ATP/ADP (adenosine diphosphate) ratio and lower concentrations of other substrates such as creatine [289]. Gene expression patterns in the hippocampus of rats fed a KD revealed an upregulation of many (but not all) transcripts encoding energy metabolism enzymes and mitochondrial proteins, compared to controls [290, 291].

Rodents fed a KD also showed increased number of hippocampal mitochondria [291, 292], a trend for increased skeletal muscle mitochondrial biogenesis (measured as citrate

synthase activity) [293], elevated glutamate levels, decreased glycogen levels and an elevated phosphocreatine:creatine ratio [291].

Concordant with these rodent studies, improved energy metabolism has been observed in humans following a KD, with increased phosphocreatine-ATP and phosphocreatine:inorganic phosphorus ratios in the grey matter compared to prior to starting the diet [294].

How exactly enhanced energy reserves may lead to improved seizure control is unclear. One interpretation is that increased ATP concentration enhances or prolongs activity of Na^+/K^+ ATPase pumps, contributing to neuronal stability [233]. Increased power for ion pumps may lead to changes in the resting membrane potential and an increased capacity to produce inhibitory neurotransmitters [165].

The 'K_{ATP}-glycolysis hypothesis' [295] is another putative explanation for the acute antiepileptic properties of the KD. According to this theory, the reduced glycolysis (and thus reduced glycolytic ATP production) brought about by low glucose/high ketone body concentrations triggers the opening of K_{ATP} channels (most importantly in the brain), depressing neuronal excitability. Although it is unknown how glycolysis is regulated in neurons in an individual following the KD, there is evidence that supports the K_{ATP}-glycolysis hypothesis. Ketone bodies (BHB or ACA) were found to reduce spontaneous firing rate in neurons in the substantia nigra pars reticulata, an effect that was seen to be even larger in faster firing cells [199]; this slowing was eliminated in the presence of K_{ATP} blockers, or in *Kir6.2* knockout mice, suggesting that K_{ATP} channels are activated/opened in the presence of ketone bodies. Consistent with this work, in the presence of BHB, basal channel and stimulus-elevated open probability were increased in mouse dentate granule neurons [203].

The authors suggested that reduced glycolytic production of ATP in the submembrane compartment could activate plasma membrane K_{ATP} channels, reducing electrical activity [199, 295]. High electrical activity, such as that produced by a seizure, increases Na^+/K^+ ATPase pump activity and ATP utilisation near the plasma membrane. This may in turn activate K_{ATP} channels in a process of negative feedback. Supporting this theory, the spike-dependent increase in K_{ATP} channel open probability in mouse dentate granule neurons was reduced or abolished in the presence of a blocker of the Na^+-K^+ ATPase [203].

Figure 1.3 depicts this negative feedback mechanism.

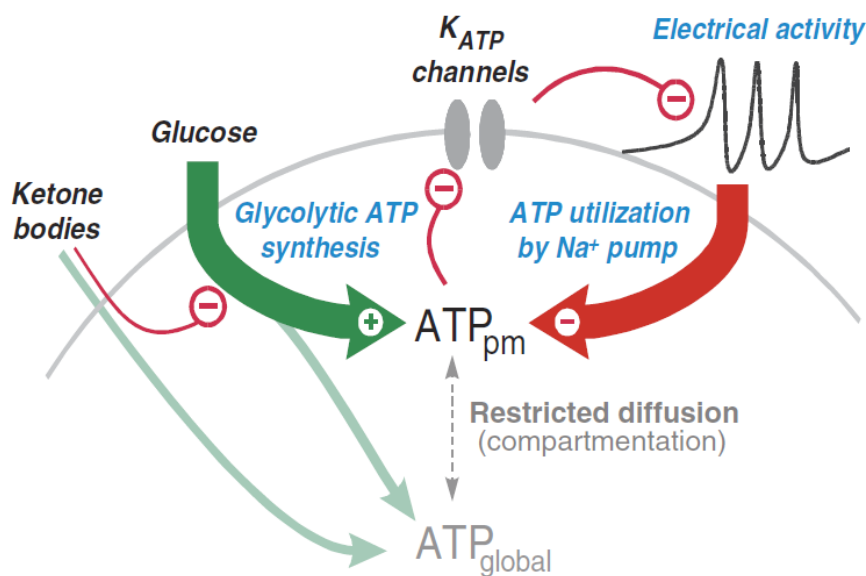


Figure 1.3: The K_{ATP} -glycolysis hypothesis, taken from Ma et al., 2007

The role of ATP in the K_{ATP} -glycolysis hypothesis may seem contradicted by observations from animal studies that the KD is associated with normal or raised hippocampal [291, 292] or whole brain [289] ATP levels [296]. Considering the role of glucose/glycolysis in seizures, it may be specifically (depleted) glycolytic energy, rather than overall increased energy reserves, that is important for seizure management, as this reduces energy reserves necessary for seizure initiation and spread [165].

Linked with this is the theory that the release of ATP associated with glucose restriction leads to elevated extracellular adenosine levels and the activation of adenosine A1R, which open K_{ATP} channels [232]. This link between activation of A1R and opening of K_{ATP} channels has been made elsewhere [232, 297-299] and its significance for the antiepileptic effect of the KD is complemented by other factors:

- a) adenosine signalling via A1R can reduce neuronal excitability [300] and suppress seizures [301].
- b) *Adk*-knockout mice, with seizures associated with deficient adenosine/A1R signalling, had no change in seizure frequency or duration when fed a KD [302]; seizures were nearly abolished in mice with intact A1R/adenosine deficiency (*Adk-Tg*) and were reduced by around 50% in mice with reduced A1R(*A1R^{+/-}*), suggesting that KD-induced seizure control depends on A1R.
- c) the presence of BHB has been associated with increased open probability of single K_{ATP} channels in dentate granule neurons [203]. Taken with findings that ketones slow neuronal firing rate via opening of K_{ATP} channels [199], this suggests that K_{ATP} channels may act as a 'link between metabolism and neuronal excitability' [203].

Evidence behind the theory of energy metabolism and the KD appears strong, and provides a plausible connection between various components of the KD and their contribution to its antiepileptic mechanism.

vi) *GABAergic Inhibition*

Alteration in brain handling of glutamate, glutamine and GABA (a major inhibitory neurotransmitter), has been proposed as another mechanism by which the KD exerts its antiepileptic effect. Data suggest that ketosis intensifies mitochondrial metabolism and flux through the tricarboxylic acid cycle, increasing the availability of glutamine and the GABA

precursor glutamate in astrocytes [303, 304]. Elevated levels of GABA are expected to dampen hyperexcitability throughout the brain [166].

In vivo and *in vitro* studies have provided evidence both for and against the GABAergic hypothesis. Rats fed a calorie-restricted KD, or a normal calorie-restricted diet, have exhibited greater paired-pulse inhibition, which may reflect enhanced GABA-ergic inhibition via enhancement of GABA_A receptors, than ad-libitum controls [187]. Calorie-restricted diets have also been associated with increased mRNA (messenger ribonucleic acid) expression of *Gad-67* and *-65* in several regions of the brain, which encode the rate-limiting enzyme in GABA synthesis, glutamic acid decarboxylase, [305]; increased *Gad67* mRNA in the striatum was an effect specific to KD-fed rats. BHB-treated astrocytes have displayed decreased GABA-transaminase and GABA transporter GAT-1 mRNA expression [306]. This has been contradicted by findings that there was no difference in regulation of transcripts encoding GABA_B receptors, metabotropic glutamate receptors, or subunits of N-methyl-D-aspartate in hippocampi from rats fed a KD, compared to controls [291].

Similar inconsistencies are found in studies regarding brain GABA levels, which may be lower [307] or unchanged [308-310] in animals fed a KD, compared to controls.

There are clinical reports that support the GABAergic hypothesis: strengthened short-latency cortical inhibition, measured by transcranial magnetic stimulation, and enhanced peri-rolandic beta activity, measured by quantitative EEG, was found in healthy volunteers following a KD for two weeks [311]. This may be a reflection of cortical GABA_A inhibition. Two of three participants with epilepsy showed increased cerebral GABA concentration, measured by magnetic resonance spectroscopy, when following a KD for two weeks [312]; one participant also showed increased GABA levels when following the KD for three months. Cerebrospinal fluid GABA was significantly increased in children with epilepsy at a

mean of four months after KD initiation, with higher levels in those who achieved >50% seizure reduction compared to non-responders [313].

vii) Norepinephrine and neuropeptides

There is evidence that norepinephrine, which protects against seizures in many animal models of epilepsy [314], is required for the antiepileptic effect of the KD. KD-fed rats had two-fold higher extracellular norepinephrine levels than rats fed a standard diet [315].

Latency to flurothyl-induced seizure was not increased in dopamine β -hydroxylase knockout mice, which are unable to synthesis norepinephrine; latency was increased in controls with normal norepinephrine levels [316].

Neuropeptide-Y and galanin, which are expressed together with norepinephrine by the majority of locus coeruleus neurons [317] and have antiepileptic properties [318, 319], are also likely candidates for the antiepileptic effects of the KD. However, the KD had no effect on the expression of genes encoding neuropeptide-Y (in the brain or liver) [192, 320] or galanin (in the brain) [320] in mice.

Leptin, which regulates the activity of these neuropeptides, has been shown to regulate hippocampal and hypothalamic neuron excitability [321] and reduce neuronal injury associated with kainic acid-induced status epilepticus [322]. Serum leptin is increased in KD-fed rats [323, 324], although hypothalamic leptin signalling may be impaired, indicated by decreased phosphorylation of STAT3 (signal transducer and activator of transcription), a downstream mediator of leptin receptor signalling. This was not due to ketones themselves, as leptin resistance was not observed with intraperitoneal injection of BHB [323].

viii) Anaplerosis

Another putative mechanism of action of the KD is based on the observation that seizures cause a deficiency in tricarboxylic acid cycle intermediates, leading to increased excitability [233, 325]. It has been hypothesised that replenishing these intermediates (a process known as ‘anaplerosis’) may prevent seizures.

The anaplerotic compound, triheptanoin, has been shown to affect seizure susceptibility in mouse models [326]. Triheptanoin supplementation of a KD reduced memory impairment in a mouse model of familial Alzheimer's disease [327].

Supplementation with the branched chain amino acids leucine, isoleucine and valine, which provide precursors to replenish tricarboxylic acid cycle intermediates [328], may increase the effectiveness of the KD [329].

This snapshot of the various hypothetical mechanisms of action of the KD highlights that a) we are still uncertain exactly how dietary treatment exerts an antiepileptic effect – there is evidence that mechanisms may be distinct/additional to those behind antiepileptic/neuroprotective effects of calorie restriction, and b) not all putative mechanisms are mutually exclusive – they may co-exist and even complement each other. This is especially true for the theory of energy metabolism, which encompasses reduced glycolysis, K_{ATP} channel modulation, enhanced adenosine transmission and improved mitochondrial function. An awareness of these (and indeed all) theories will influence decisions regarding genetic analyses, including the selection of candidate genes, and potentially the interpretation of results.

1.2.6 Are we able to predict response to the Ketogenic diet?

The KD is the treatment of choice for GLUT1 deficiency syndrome [330] and PDD [40, 331, 332], as it provides an alternative fuel source, relieving blocks in metabolism upstream

from the tricarboxylic acid cycle [333]. The KD has also been found to be particularly useful for the following conditions/syndromes:

- epilepsies with structural causes, such as lissencephaly, hypoxic-ischemic encephalopathy, cerebrovascular accident, intraventricular hemorrhage [334] or focal malformation of cortical development [96].
- status epilepticus [335] and epilepsies with immune causes, such as fever-induced refractory epileptic encephalopathy in school age children [336, 337].
- certain epilepsy syndromes, often with genetic causes, such as epilepsy with myoclonic atonic seizures [338-340], infantile spasms [57, 115, 341, 342], mitochondrial respiratory chain defects [343], Lennox-Gastaut [82, 113] and Dravet syndrome [79, 344-347]. It is of interest that 5% of a cohort of children with epilepsy with myoclonic atonic seizures were found to have GLUT1 deficiency with mutations in *SLC2A1* [348]. Increasing numbers of patients with early-onset absence epilepsy are being found to have *SLC2A1* mutations [349, 350]; these individuals may not necessarily be drug-resistant or have developed a movement disorder.

This list acts merely as an indication of likelihood of KD response; it does not rule out a favourable response to the KD in patients with other epilepsy syndromes or epilepsies of different/unknown causes.

The question remains whether other patient demographics or clinical parameters allow response to the KD to be predicted. Such an exploration of potential predictors of response is essential for this study, as it may allow potential confounding factors in this cohort to be identified and enhance understanding of the mechanisms behind dietary therapy for epilepsy.

A PubMed search of the English- and Spanish-language literature was performed, with the keywords 'ketogenic diet' and 'epilepsy'. Articles regarding the KD for metabolic disorders and refractory status epilepticus were excluded, as were case reports, abstracts, letters and papers published pre-1970.

No single factor was consistently shown to predict response (favourable or unfavourable) to the KD, as outlined in a recent review [351].

- i) The following factors have (almost) consistently demonstrated no effect on KD response:

Gender: gender has been found to have no effect on response to the KD in children [57, 66, 71, 75, 79, 86, 93, 105, 107, 113, 118, 119, 123, 124, 127, 208, 211, 213, 279, 339, 344, 352-360], adolescents [76, 93, 113, 118, 119, 123, 211, 279, 344, 352, 353, 355, 356, 359, 360] or adults [100, 121, 355, 360].

Others have found that male children and adolescents were slightly more likely to respond to the KD after one month, compared to females [55]. In contrast, female gender has predicted a favourable response after 12 months in adults [126].

Intellectual status: developmental delay and cognitive impairment have been found not to influence KD effectiveness [72, 76, 103, 208, 354, 361], although individuals with less severe mental delay who have achieved seizure reduction on the KD may be more likely to maintain this reduction at the 12-month point [361].

- ii) Evidence regarding the effects of other factors on KD response is mixed:

Age at diet initiation: no difference has been found when comparing age at diet initiation with KD response in children [54, 55, 66, 71, 79, 84, 86, 90, 93, 96, 105, 107, 113, 118, 119,

124, 127, 208, 211, 213, 279, 339, 347, 354-360, 362-364], adolescents [54, 55, 76, 93, 113, 118, 119, 279, 355, 356, 359, 360, 363] or adults [100, 121, 126, 355, 360].

Others have found KD response to be more favourable in children under one year of age [75], under 3 [365], under 5 [366], under 8 [352], under 10 [367], under 12 years [72], and with a 'lower average age' [123] compared to older children in each cohort, and in those with a mean age of 8 compared to non-responders with a mean age of 12 years [37].

Improved response has also been found in those aged between 3 or 4 - 10 years, compared to toddlers and older children [130, 353].

Age at seizure onset: age at seizure onset has been found not to affect response to the KD in children [54, 55, 57, 86, 107, 109, 118, 124, 127, 208, 344, 360, 362], adolescents [54, 55, 118, 344, 360] or adults [121, 360].

Others have found an association between lower age at epilepsy onset with a favourable response to the KD in children [211, 364] and adolescents [211]. Higher age at onset of infantile spasms has been associated with increased likelihood of spasm reduction with dietary therapy [105].

Anti-epileptic drugs: no correlation has been found between KD response and number of AEDs previously tried [55, 118, 121, 123, 124, 127, 208, 360], or number of AEDs exposed to at the time of diet initiation [55, 107, 118, 121, 126, 127, 355, 360, 362]. Differences in AEDs used or drug combinations [211], changes in AEDs [93], early/late reduction of medication during dietary treatment [78], or time to initial treatment of spasms with medication [109] may also not affect diet response.

Current or previous use of valproic acid seems not to affect KD response [83, 113].

Others have found a favourable response to the KD in children exposed to fewer AEDs prior to diet initiation [75, 105] and at diet initiation [105], compared to non-responders. A higher likelihood of response may also be found in children with no prior steroid use [105].

Children receiving phenobarbital in conjunction with the KD have been found to be less likely to respond to dietary treatment than those receiving valproic acid, topiramate, levetiracetam, lamotrigine or zonisamide; those receiving zonisamide may be more likely to respond [368].

Body Mass Index (BMI): no relationship has been found between KD response and BMI [118, 127, 208, 358] or body weight [121] at diet initiation. Response to the KD may also be independent of the following factors during dietary treatment: BMI changes [93, 128, 358]; weight loss [55, 126], reduction in BMI [126]; growth [369].

Others have found an association between increased likelihood of KD response with a lower weight-to-height z-score [109], a stable BMI [118] and with a greater decrease in BMI during dietary treatment [121].

Diet ratio: no difference has been found in KD response with different ketogenic ratios [57, 75, 76, 83, 105, 123, 208, 370].

Individuals following a KD with a 4:1 ketogenic ratio had better seizure control than those following a 3:1 diet [51]. People who started a diet on 10g carbohydrate/day were found to be more likely to respond to treatment at 3-month follow-up (but not at other follow-up times) than those who started on 20g/day [50].

Diet type: a 'strong trend' has been identified for higher responder rates in people following the classical KD compared to those following the MAD [125] and a higher proportion of children following the classical KD reported decreased (unknown to what

extent) seizure frequency and severity, compared to those following the MAD [371]; others have not found this comparing the classical KD with the MAD [101, 360], or with the MCT KD [98, 278, 279].

EEG: no relationship has been reported between KD response and EEG patterns in general [213, 358, 364], or any of the following parameters: classic hypsarrhythmia, focal versus non-focal EEG abnormalities, EEG slowing, spikes or polyspikes, presence of focal spikes or EEG changes, high voltage spike-wave or multifocal discharges, normal background, epileptiform discharges in frontal, central, parietal and occipital regions, bilateral synchrony, trains of epileptiform discharges, burst suppression patterns, or generalised spike-wave abnormalities, posterior delta/theta/alpha relative power, spike index, number of regions of epileptiform discharges, or effect of sleep on epileptiform discharges [57, 66, 71, 103, 105, 107, 124, 208].

Others have found an unfavourable outcome to the KD to be associated with presence of epileptiform EEG discharges in the temporal region [103], electrographic bilateral synchrony [103], higher frequency EEG background activity [211], classic hypsarrhythmic patterns [75] and multifocal spikes [66] at diet onset. No authors specifically stated whether this was independent of epilepsy syndrome or not.

Epilepsy cause or syndrome: of those studies that specifically looked at KD response in people with epilepsies of different causes or with different epilepsy syndromes, the majority found no difference between groups: diet response may be unaffected by epilepsy cause [57, 72, 75, 80, 84, 90, 103, 113, 119, 124, 211, 356-358, 363, 372], epilepsy syndrome [119, 211, 355, 357, 358, 362], or by prior resective surgery [121, 126]. A favourable response to the KD has been found to be more likely in people without 'surgically-approachable epilepsy' or a surgically-approachable focus [95, 133, 373] (this

may be a surrogate for epilepsy syndrome), although others have reported high response rates in individuals with surgically-remediable focal malformations of cortical development [96].

Others have reported a more favourable response to the KD in individuals with symptomatic, as opposed to idiopathic or cryptogenic epilepsy [354, 355]. Better response rates have also been reported in those with 'idiopathic' or 'cryptogenic' epilepsy [84, 86].

A higher proportion of children with infantile spasms caused by pre/perinatal injury were found to respond to the KD, compared to those with infantile spasms caused by brain malformation or with cryptogenic spasms [109].

Seizure-freedom whilst following the KD has been achieved in more people with epileptic encephalopathies than in those with localisation-related epilepsies [374].

Children with myoclonic-astatic epilepsy were found to be more likely to respond to the KD than those with absence epilepsy, cortical dysplasia, Lennox-Gastaut syndrome, Sturge-Weber syndrome, myoclonic absence epilepsy, juvenile myoclonic epilepsy, encephalitis, Angelman syndrome or Dravet syndrome [127].

Ketosis: no correlation has been found between seizure control and blood BHB levels in children [51, 55, 56, 109, 119, 128, 210-213, 375], adolescents [51, 55, 56, 119, 128, 210-212, 375] or adults [126, 128, 210, 375], nor between seizure control and urinary ketosis in children [50, 55, 80, 93, 122-124, 132, 209, 375], adolescents [50, 55, 93, 122-124, 132, 209, 375] or adults [100, 121, 132, 209, 375].

'Large ketosis' (unspecified whether ketones measured in blood or urine) has been found not to influence the likelihood of achieving seizure-freedom during the first two weeks of following the KD [208].

No relationship has been found between breath acetone and seizure control in children, adolescents and adults following the KD [376].

Others have reported a correlation between improved seizure control and higher blood BHB levels in children [36, 98, 132, 193], adolescents [98, 132] and adults [132] following the KD.

A positive correlation has been found between seizure control and urinary ketone levels in children, [98, 118, 119, 354, 369], adolescents [98, 118, 119, 369] and adults [369].

Other biochemical markers: KD responders have been found to have larger absolute decreases in plasma phospholipid fatty acid 18:0 and lesser increases in 24:1 during dietary treatment [89] compared to non-responders. Positive correlations between reduction in seizure frequency and elevation in serum total arachidonic acid [212], increase in fat oxidation [377], increase in serum palmitoleic acid and decrease of serum arachidonic acid [101] have also been reported.

In one cohort, the individual who achieved greatest seizure reduction on the KD also had the largest increase of total and low-density lipoprotein (LDL) cholesterol [120].

No difference has been found in levels of the plasma phospholipid fatty acids 14:0, 16:0, 16:1, 18.1, 18.2, 18.3, 20.0, 20.3, 20.4, 20.5, 22.0, 22.5, 22.6 and 24.0 (number of carbons: number of double bonds) [89], monoamine or HVA/5HIAA levels [378] between diet responders and non-responders, measured pre-diet and during dietary treatment.

Responders and non-responders have displayed no difference in levels of carnitine, amino acids, organic acids or blood count [211], levels of linoleic acid or alpha-linolenic acid [101], or in levels of octanoic or decanoic acid [276].

Seizure frequency: pre-diet seizure frequency has been found not to affect KD response [55, 75, 76, 100, 103, 105, 118, 126, 208, 352, 360].

Others have reported KD response to be more favourable in people with recently increased seizure frequency [54] or a higher pre-diet seizure frequency [121, 127].

Seizure type: a more favourable response to the KD has been reported in people with generalised seizures compared to partial seizures [72, 90, 95, 211, 366, 374].

Focal seizures have been associated with decreased likelihood of a favourable response to the KD compared to other seizure types [50, 93, 103, 208]. In one cohort, dietary therapy was ineffective for 'grand mal or focal motor seizures' [37]. In another, children with 'minor seizures' showed a better response to the KD compared to those with 'major seizures' [379].

Children with infantile spasms [354] or multiple seizure types [76, 93] have been found to respond more favourably to the KD than those with other specific seizure types.

Others have found no differences in seizure control when comparing seizure types in people following the KD [55, 66, 71, 79, 80, 84, 85, 107, 113, 119, 121, 124, 128, 145, 153, 213, 279, 339, 344, 352, 354, 355, 358, 359, 361, 363, 365, 380].

Time between seizure onset and initiation of dietary treatment: no difference has been found in response to dietary treatment with respect to time between epilepsy onset and dietary treatment [54, 57, 86, 100, 103, 109, 213].

Others have shown KD responders to have a shorter time from epilepsy onset to diet initiation than non-responders [99, 211].

- iii) There are limited data (\leq four studies per factor) regarding the effects of the following factors on KD response:

Blood glucose: a correlation has been reported between lower blood glucose levels and increased diet effectiveness [128, 229]. Others, however, have found that blood glucose levels were not correlated with diet response [109, 230].

Imaging: no difference in likelihood of achieving seizure freedom within two weeks of diet onset has been reported between people with or without a structural brain abnormality [208]. KD effectiveness may not be predicted by magnetic resonance imaging [107, 124] or computerised tomography findings [124].

Other: KD response has been found to be independent of ethnicity [93], having a family history of epilepsy or history of febrile seizure [103].

In summary, there is no evidence to suggest that KD response is affected by any of the factors identified. For the majority of factors, findings are inconsistent, which makes it difficult to draw conclusions. Some factors, such as differential response to the diet with age, may also be affected by environmental and/or social factors. Even the apparently most consistent evidence must be interpreted with caution. For example, only one of the studies that investigated the relationship between ketosis and effectiveness of the KD at various follow-up points found this correlation to be true at all times [159]; this may indicate that ketosis serves more as an indicator of dietary compliance, rather than effectiveness.

Reports of improved response in very young children and adolescents may be due to better diet adherence in these age groups, and small numbers of participants in some age groups must be considered. Disease severity, referral bias and resource availability may also affect age at KD initiation.

It must also be considered that each factor may not be a separate entity. For example, the following factors may be inextricable from epilepsy syndrome, or at least represent a greater or lesser degree of epilepsy severity: age at diet initiation, age at seizure onset, EEG

and imaging findings, intellectual status, pre-diet seizure frequency and surgery or surgically approachable epilepsy. Brain ketone utilisation depends on other factors, including blood ketone and glucose levels [221, 381], which may explain the variability in study findings with respect to the effect of these factors on KD response. Reports of improved response to the KD in people with more severe or recently-worsened epilepsy may simply reflect a regression to the mean.

Evidence suggesting no effect on KD response is most consistent for gender and intellectual status. A large number of studies found no effect of gender on diet response and the two studies that do report a gender effect, which comprise different age groups and follow-up times, are conflicting. No studies have reported an effect of intellectual status on KD response; the finding that individuals with less severe mental delay who achieved seizure reduction on the KD were more likely to maintain this reduction at long-term follow-up may be due to increased likelihood of compliance with the diet, or may reflect epilepsy type.

In the absence of sufficient or consistent evidence from human studies, *in vivo* and *in vitro* models may provide further clues.

Many factors and their putative effect on KD response have been investigated in experimental studies; a variety of models have been used, so collated results must be interpreted with caution.

The KD can be effective in both young and adult rodents [166], although some have shown a calorie-restricted KD to be most effective in young rats [382].

BHB levels have been shown to correlate [227, 383] and not correlate [186, 194, 214-219] with KD response in rodents.

Likewise, blood glucose levels have been shown to correlate [185, 228] and not correlate [191, 216, 217, 219, 227] with KD response in rodents.

Seizure protection in rats may not be affected by fat composition of a KD [215], although PUFA supplementation of a standard diet can exert anticonvulsive properties comparable to the KD [243].

Studies have shown increased diet ratio to be associated with improved KD effectiveness [214, 217, 227, 384] in rodents.

A calorie-restricted KD was found to induce a significant increase in seizure resistance in rats compared to an ad libitum KD [385].

One study found that valproate, carbamazepine, lamotrigine, oxcarbazepine, and topiramate significantly decreased the concentration of free acetone in mouse brain [386].

Experimental observations regarding the effects of certain factors on KD response tend to replicate their clinical correlates. For example, animal studies provide conflicting evidence with regards age at diet initiation and blood glucose. Experimental evidence indicating improved diet effectiveness with increased diet ratio does appear strong, which is inconsistent with most clinical findings. This may be due to the fact that the ketogenic ratios often used in animal studies are higher than those used (or that could be used in a palatable diet) in humans. Most animal studies have demonstrated a lack of correlation between ketosis and seizure protection, which may point towards an indirect effect of ketone bodies in the KD, such as optimising Krebs cycle function or inhibition of ROS production [296]; ketone body levels may be a surrogate for the real mechanism of the KD [273].

Overall, the limited or conflicting evidence has not highlighted any specific patient demographics or clinical parameters to be mindful of in future analyses. It also provides little indication as to potential mechanisms of action of the KD in humans. This literature review highlights the need to look for predictors of response to the KD, as (with the exception of specific metabolic disorders) there is currently no consistent evidence for any one predictor. The rationale behind investigating potential genetic predictors of response will be explored in the following section.

1.3 The Ketogenic diet and Genetics

In the context of disease and pharmacotherapy (or perhaps, more appropriately in this case, ‘nutritherapy’), a comprehensive understanding of the interaction of genotype and phenotype holds immense potential. For example, the clinical utility of the discovery of the association of carbamazepine-induced Stevens-Johnson syndrome (SJS) and toxic epidermal necrolysis (TENS) with the HLA-B*1502 allele in Asians [387] has been demonstrated: prospectively screening Han Chinese individuals before initiating carbamazepine therapy, and withholding carbamazepine from those who were HLA-B*1502–positive, reduced the incidence of carbamazepine-induced SJS/TENS [388]. Pre-treatment screening may theoretically reduce the financial burden of drug-related adverse outcomes [389], but this has not been shown to be the case with HLA-B*1502 screening for people with epilepsy [390].

Differences in gene expression may also be used as phenotypic predictors, such as drug-response in cancer [391], weight maintenance following calorie restriction [392] and response to endurance exercise training [393].

It seems prudent to investigate whether genetic variation influences KD response due to various factors. A particularly favourable response to the KD has been observed in

individuals with conditions mainly caused by single genetic factors and there are limited animal and human studies that suggest a genetic basis to KD response. As a period of time (up to some weeks) tends to be needed for the KD to exert its full anticonvulsant effects, it has been suggested that altered gene expression is involved [394]; the KD has been associated with gene expression changes in animal models, as will be discussed in the following section. One of the major influences on gene expression is gene variation [395-397] and so it seems appropriate that individual genetic variation determines whether somebody responds favourably to the KD or not.

An awareness of the literature associated with the KD, its putative antiepileptic mechanisms of action and the theoretical interaction of genetics has facilitated the selection of candidate genes for further study. Further detail will be given in Chapter 4.

1.3.1 Differential response to the Ketogenic diet

There is evidence that certain epilepsy syndromes and conditions in humans respond particularly well to the KD [151, 398]. Many of these conditions are caused predominantly by single genetic factors, which may lead to the postulate that there are other genetic factors, independent of epilepsy syndrome, that also influence response to the KD.

Seven epilepsy syndromes have a 'probable benefit' (reported in at least two publications) from the diet [151]; these syndromes, along with their causative mutations, are outlined in Table 1.4. Several syndromes, as well as selected mitochondrial disorders, such as phosphofructokinase deficiency and mitochondrial respiratory chain complex disorders, have a 'suggestion of benefit' (only from case reports or series); these can be found in Table 1.5. Associated genes were found from Online Mendelian Inheritance in Man (<http://www.omim.org>), Genetics Home Reference (<http://ghr.nlm.nih.gov>), The MalaCards human disease database (<http://www.malacards.org>), and PubMed searches.

Table 1.4: Conditions caused by genetic factors that have a ‘probable’ benefit from the Ketogenic diet

CONDITION	ASSOCIATED GENE MUTATIONS	
Glucose transporter protein 1 (GLUT1) deficiency	<i>SLC2A1</i>	
Pyruvate dehydrogenase deficiency	<i>PDHA1</i> <i>PDHB</i> <i>DLAT</i> <i>DLD</i> <i>PDHX</i> <i>PDP1</i>	
Epilepsy with myoclonic atonic seizures (Doose syndrome)	<i>SCN1A</i> <i>SCN1B</i> <i>GABRG2</i> <i>GRIN2B</i> <i>SLC2A1</i>	
Tuberous sclerosis complex	<i>TSC1</i> <i>TSC2</i>	
Rett syndrome	<i>MECP2</i> <i>CDKL5</i> <i>FOXP1</i>	
Severe myoclonic epilepsy in infancy (Dravet syndrome)	<i>SCN1A</i> <i>SCN1B</i> <i>SCN2A</i> <i>SCN9A</i> <i>PCDH19</i> <i>GABRG2</i> <i>GABRD</i>	
West syndrome	<i>CDKL5</i> <i>ARX</i> <i>SCN1A</i> <i>SCN2A</i> <i>MAGI2</i> <i>LINC00581</i> <i>FOXP1</i> <i>SPTAN1</i> <i>ALDH7A1</i> <i>GLDC</i> <i>GCSH</i> <i>GCST</i> <i>KCNJ11</i> <i>POMC</i> <i>PNKP</i> <i>FLNA</i>	<i>STXBP1</i> <i>IL1RN</i> <i>PHGDH</i> <i>TLE</i> <i>ARX</i> <i>PTCH1</i> <i>TSC1</i> <i>TSC2</i> <i>LIS1</i> <i>DCX</i> <i>SLC25A22</i> <i>ST3GAL3</i> <i>TUBA1A</i> <i>ATP7A</i> <i>TPH1</i> <i>IDUA</i>

Table 1.5: Conditions caused by genetic factors that have a 'suggested' benefit from the Ketogenic diet

CONDITION	ASSOCIATED GENE MUTATIONS	
Phosphofructokinase deficiency (Glycogen storage disease type VII)	<i>PFKM</i>	
Myophosphorylase deficiency (Glycogen storage disease type V)	<i>PYGM</i>	
Mitochondrial respiratory chain complex I	<i>ACAD9</i>	<i>NDUFAF3</i>
	<i>FOXRED1</i>	<i>NDUFAF4</i>
	<i>MT-ND1</i>	<i>NDUFAF5</i>
	<i>MT-ND2</i>	<i>NDUFAF6</i>
	<i>MT-ND3</i>	<i>NDUFS1</i>
	<i>MT-ND4</i>	<i>NDUFS2</i>
	<i>MT-ND4L</i>	<i>NDUFS3</i>
	<i>MT-ND5</i>	<i>NDUFS4</i>
	<i>MT-ND6</i>	<i>NDUFS6</i>
	<i>NDUFA1</i>	<i>NDUFS7</i>
	<i>NDUFA10</i>	<i>NDUFS8</i>
	<i>NDUFA11</i>	<i>NDUFV1</i>
	<i>NDUFA2</i>	<i>NDUFV2</i>
	<i>NDUFAF1</i>	<i>NUBPL</i>
	<i>NDUFAF2</i>	<i>NDUFB3</i>
Mitochondrial respiratory chain complex II	<i>SDHAF1</i>	
	<i>SDHA</i>	
Mitochondrial respiratory chain complex IV (Cytochrome-c Oxidase Deficiency Disease)	<i>FASTKD2</i>	<i>MTCO1</i>
	<i>COX14</i>	<i>MTCO2</i>
	<i>COX6B1</i>	<i>MTCO3</i>
	<i>COA5</i>	<i>COX15</i>
	<i>MTTS1</i>	<i>COX10</i>
	<i>MTTL1</i>	<i>SURF1</i>
	<i>C2ORF64</i>	<i>SCO2</i>
	<i>C12ORF62</i>	<i>SCO1</i>
	<i>COX20</i>	<i>TACO1</i>
	<i>MTTN</i>	<i>LRPPRC</i>
Lafora body disease	<i>EPM2A</i>	
	<i>NHLRC1</i>	
	<i>NFU1</i>	
	<i>CSTB</i>	
	<i>CLN3</i>	
	<i>GBE1</i>	
	<i>PARK2</i>	
	<i>PPP1R3C</i>	
	<i>TRIM32</i>	
	<i>PRDM8</i>	

Few studies have reported on KD response in individuals screened for genetic mutations.

2/4 people with *SCN1A* mutations and 1/1 with a 4p- deletion achieved >50% seizure

reduction after following a KD for three months [334]. When following the diet for one

month, 8/13 (61.5%) children with *SCN1A* mutations achieved $\geq 75\%$ seizure reduction; at 3- and 6-month follow-up, 6/13 (46.2%) achieved $\geq 75\%$ seizure reduction, as did 5/13 (38.5%) at 9-month follow-up [345]. Another study tested individuals with Dravet syndrome for *SCN1A* mutations (6/7 tested positive) and reported on the effectiveness of dietary treatment in six individuals (4/6 had improved seizure control), but it is unclear whether the KD was administered in those tested for mutations or not [346].

No studies have directly compared KD response with presence/absence of specific genetic mutations in humans.

Animal studies have shown strain-specific responsiveness to the KD in terms of seizure threshold: FVB/NJ in-bred mice strains fed a KD showed increased latency to generalised tonic clonic seizure, compared to controls; A/J mice showed reduced latency [399].

The KD has been shown to reduce seizure frequency in *Kcna1*-null mice [400] and elevate seizure threshold in *Scn1a* knockout mice [401].

The KD resulted in near seizure-freedom in *Adk-Tg* mice, 'significantly' reduced seizures in *A₁R^{+/-}* mice, and had no effect in *A₁R^{-/-}* mice, suggesting a role of adenosine A₁R in KD-induced seizure reduction [302].

Genetic influences on biochemical parameters associated with the KD have been reported.

Increased mitochondrial utilisation of BHB was found in *Bad*-null and *Bad^{S155A}* cortical neurons and astrocytes compared with wild-type cells, leading to increased open probability of K_{ATP} channels [226] (this will be discussed further in Chapter 4); higher activity levels of hepatic BHB dehydrogenase and acetoacetyl-CoA thiolase were found in *Shc* knockout compared to wild-type mice [402]. KD- or fasting-induced ketosis was suppressed in *Prkce*-null [403], *Ppara*-null [404] and *Fgf21* knockout [405] mice compared to wild types, although evidence is conflicting for the latter [406]. Glucose levels have been found to be

lower in KD-fed *Ppara*-null mice [404] and higher in *Fgf21* knockout mice [405], compared to wild types.

Differential responsiveness to the KD with a genetic basis has also been shown in patients with Alzheimer's disease. Daily administration of the ketogenic agent AC-1202 for 90 days resulted in significant differences in serum BHB levels and Alzheimer's Disease Assessment Scale-Cognitive subscale (ADAS-cog) scores compared to placebo; effects were most notable in patients who did not carry the apolipoprotein E4 (*APOE4*) allele [407].

Consumption of an MCT drink, compared to placebo, has led to improved performance in ADAS-cog in *APOE4*- but not *APOE4*+ subjects [408].

Experimental work indicates that the KD is effective in *Kcna1*- and *Scn1a*-null mice and that responsiveness to the KD (in terms of seizure control) may be strain-specific, but further studies are needed. The following genes have been shown to play a potentially important role in the mechanisms of the KD: *Bad*, *Kcnj11*, *Shc1*, *Pkce*, *Ppara* and *Fgf21*, but it is unknown whether this is relevant for humans. Genetic factors may influence efficacy of the KD, in terms of improved cognition, in humans.

1.3.2 Differential regulation of gene expression

Decades of research have led to the identification of many gene expression changes associated with calorie restriction [409]; one would expect similar changes from the KD. Indeed, many of the putative mechanisms of action of the KD involve an up- or down-regulation of gene expression.

A recent study showed that pilocarpine-injected, epileptic rats fed a KD had a distinct genomic methylation profile and gene expression patterns compared to untreated epileptic

rats – the KD seemed to ameliorate DNA methylation and subsequent gene expression changes associated with seizures [410]. A full list of differentially methylated or expressed genes was not provided. Only two candidate genes were mentioned: KD treatment attenuated the hypermethylation and decreased gene expression of *Camkk2* (calcium/calmodulin-dependent protein kinase kinase 2, beta) in the epileptic rat (*Camkk2* gene expression was decreased in non-KD-fed epileptic rats compared to controls), a difference which was statistically significant; the KD reversed the hypomethylation of the *Il10rb* (interleukin 10 receptor, beta) locus seen in non-KD-fed epileptic rats but it had no effect on *Il10rb* expression.

A literature search was conducted to identify other genes that have been shown to be differentially regulated in animal models (both ‘healthy’ and epilepsy models were included) due to the KD. Only genes shown to be upregulated from feeding of a KD, not from the application of substances designed to mimic the effects of KD-feeding, such as 2DG or BHB, were considered. This is to allow for a more accurate representation of KD-feeding in humans, and allows for the (presumed) interaction of the many components of the KD.

Figure 1.4 portrays a snapshot of genes, and the proteins they encode, that have been shown to be upregulated by the KD in animal models. Official full names and gene symbols are taken from the *Gene* database (www.ncbi.nlm.nih.gov/gene), searched for by Accession number or by the gene name used by study authors. Mouse Genome Informatics nomenclature guidelines have been followed (*Guidelines for Nomenclature of Genes, Genetic Markers, Alleles, and Mutations in Mouse and Rat*: <http://www.informatics.jax.org/mgihome/nomen/gene.shtml>).

The potential consequences of each gene upregulation are given in yellow boxes, where possible.

Other genes shown to be upregulated by the KD include the following:

Latexin (*Lxn*) [290]

TIMP metalloproteinase inhibitor 2 (*Timp2*) [290]

Bone morphogenetic protein receptor, type IA (*Bmpr1a*) [290]

Nuclear receptor subfamily 3, group C, member 2 (*Nr3c2*) [290]

Guanine nucleotide binding protein (G protein), alpha z polypeptide (*Gnaz*) [290]

Myelin basic protein (*Mbp*) [290]

Syntaxin 1B (*Stx1b*) [290]

Synaptotagmin XI (*Syt11*) [290]

Prion protein (*Prnp*) Θ [283]

Membrane protein, palmitoylated 4 (*Mpp4*) Θ [283]

Neuroglobin (*Ngb*) Θ [283]

Serine (or cysteine) peptidase inhibitor, clade E, member 1 (*Serpine1*) \mathbb{X} [404]

Glucose transporter 1 (*Glut1*) Θ [411]

Glucose transporter 3 (*Glut3*) Θ [411]

and 389 genes listed in Supplementary Table 1 in Bough et al., 2006 [291].

Θ =shown in whole brain

\mathbb{X} =shown in liver

The effects of these genes with regards to KD antiepileptic action have not been extensively investigated.

Figure 1.5 portrays a snapshot of genes (and the proteins they encode) that have been shown to be downregulated by the KD in animal models. The same criteria used for Figure 1.4 have been applied.

KEY (for Figures 4 and 5)

Black text with no symbol=shown in hippocampus

Θ =shown in whole brain

\mathbb{X} =shown in hippocampus and liver

\mathbb{S} =shown in hippocampus and whole brain

\mathbb{X} =shown in liver

\mathbb{T} =shown in lymph node and CNS

\mathbb{S} =shown in striatum

*Energy metabolism genes

Genes from Fig 3B Bough et al., 2006 [291] (Appendix 1.1)

ATP synthase, H⁺ transporting, mitochondrial Fo complex, subunit d (*Atp5h*)

ATP synthase, H⁺ transporting, mitochondrial Fo complex, subunit B1 (*Atp5f1*)

Acyl-CoA synthetase long-chain family member 1 (*Acs1l*)

aldolase A, fructose-

bisphosphate (*Aldoa*)

NAD(P)H dehydrogenase, quinone (*Nqo1*) ⊖

**Mitochondrial transcripts:

See Appendix 1.2: transcripts that encode mitochondrial proteins, from Bough et al., 2006 [291]

***Peroxisome proliferator-activated α [192, 256, 283, 289-291, 305, 412-414] and δ have not been shown directly to be upregulated by KD-feeding

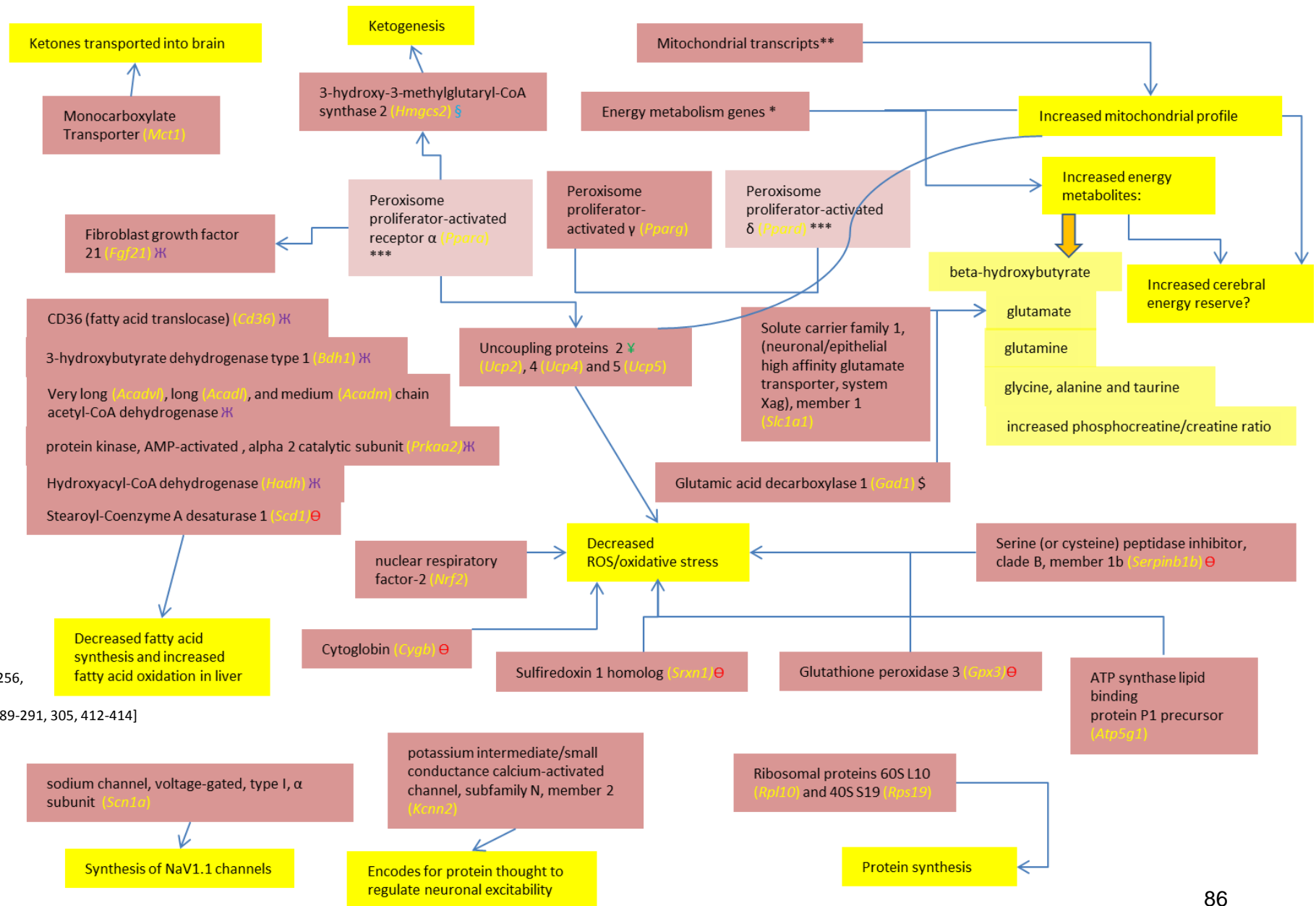


Figure 1.4: Genes shown to be upregulated by the Ketogenic diet in animal models

[192, 290,
291, 405,
413, 415-
418]

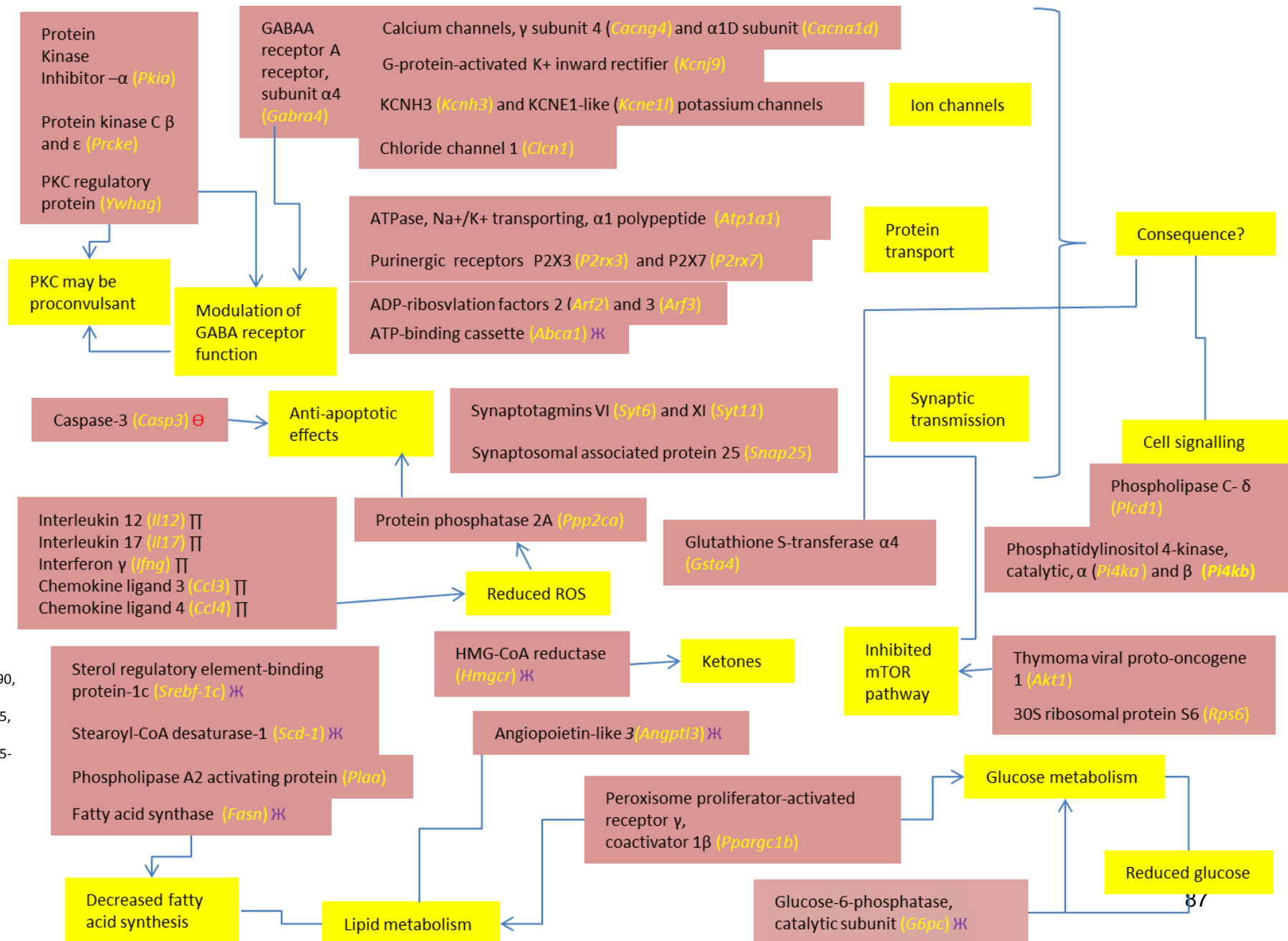


Figure 1.5: Genes shown to be downregulated by the Ketogenic diet in animal models

Other genes shown to be downregulated by the KD:

Cytochrome P450, family 17, subfamily a, polypeptide 1 (*Cyp17a1*) [290]
Macrophage migration inhibitory factor (*Mif*) [290]
Tachykinin 2 (*Tac2*) [290]
Proopiomelanocortin (*Pomc*) \emptyset [192]
Ras-related protein, (*Rala*) [290]
Calponin 3 (*Cnn3*) [290]
Fibrinogen β chain (*Fgb*) [290]
V-erb-b2 erythroblastic leukemia viral oncogene homolog 3 (*ErbB3*) [290]
Casein kinase II β subunit (*Csnk2b*) [290]
Apolipoprotein E (*ApoE*) Θ [415]
Amyloid beta (A4) precursor protein (*App*) Θ [415]
Heat shock protein 90 α (*Hsp90ab1*) [290]
and 274 genes listed in Supplementary Table 1 in Bough et al., 2006 [291].
 \emptyset =shown in hypothalamus
 Θ =shown in whole brain

Figure 1.4 and Figure 1.5 emphasise the complexity of KD treatment and the fact that many of the gene expression changes have unknown consequences.

In summary, the principal effects on gene expression induced by the KD are:

In the hippocampus:

- Upregulation of mitochondrial transcripts and genes associated with energy metabolism. This has led to an increased number of mitochondrial profiles and energy metabolites (although evidence is conflicting as to which metabolites are increased).
- Upregulation of PPARs. Although only *Ppar γ* has been shown to be upregulated with the KD (with effects on UCPs), *Ppara α* and *Ppara δ* are also assumed to be implicated [256].
- Upregulation of *Mct1*, leading to increased transport of ketones into the brain.
- Upregulation of genes encoding subunits of sodium and potassium channels. One would expect this to affect regulation of neuronal excitability and seizure susceptibility

[304] (many genes associated with epilepsy encode voltage-gated sodium channels [419] and *Kcnn2* has been found to protect against kainate and glutamate excitotoxicity and increase neuron survival after sodium cyanide or staurosporine insult [420]) although this has not been shown directly by KD-feeding.

- Differential regulation of genes leading to reduced ROS/oxidative stress.
- Downregulation of protein kinase C-related genes. One would expect this to affect neuronal excitability, potentially through GABA receptor activation – there is evidence that the receptor for activated C kinase (*Rack1*) and protein kinase C bind to distinct sites on the GABA_A receptor, potentially affecting neuronal excitability [421].
- Downregulation of genes encoding ion channels, and genes associated with protein transport, synaptic transmission, and cell signalling. How exactly this may exert an antiepileptic effect is unknown.
- Downregulation of *Ppp2ca*, possibly due to reduced oxidative stress, leading to reduced apoptosis through the inactivation of BCL2 [422].
- Downregulation of *Rps6* and *Akt1*, potentially leading to an antiepileptic effect via inhibition of the mTOR pathway.

In whole brain:

- Upregulation of *Hmgcs2*, possibly via *Ppara*.
- Upregulation of other genes associated with reduced ROS/oxidative stress.
- Downregulation of *Casp3*, leading to antiapoptotic effects.

In the liver:

- Upregulation of *Hmgcs2*, possibly via *Ppara*, and downregulation of *Hmgcr*, leading to increased serum ketone bodies. Fasting/drug-activated *Ppara* upregulates expression of

many other genes, such as medium chain acyl dehydrogenase (*Mcad*), carnitine palmitoyltransferase 1a (*Cpt1a*) and long chain fatty acyl-CoA synthase (*Acs1*) [256]; these have not directly been shown to be affected by the KD.

- Upregulation of *Fgf21*. PPAR α -mediated expression of FGF21 in the liver stimulates gluconeogenesis, fatty acid oxidation and ketogenesis [423]. It is unknown how this may affect seizure susceptibility, but it is of interest that FGF21 can cross the blood brain barrier [424] and that rats treated with intracerebroventricular FGF21 had suppressed hepatic glucose production and gluconeogenic gene expression, and increased insulin sensitivity [425].
- Differential regulation of genes associated with fatty acid metabolism, consistent with increased fatty acid oxidation and decreased fatty acid synthesis.
- Downregulation of *G6pc*, leading to decreased gluconeogenesis.

In the lymph node and central nervous system (CNS):

- Downregulation of genes encoding cytokines and chemokines, potentially leading to ROS reduction.

In the striatum:

- Increased expression of *Gad1*. This may be associated with increased GABA synthesis.

There is conflicting evidence regarding the regulation of other genes from KD-feeding:

Tumor necrosis factor α (Tnfa): Hippocampal expression *Tnfa* has been shown to be increased in KD-fed mice, compared to controls, and kainic acid-induced increases in *Tnfa* were attenuated by KD-feeding [414]. Others have found no change in hepatic *Tnfa* expression [405], and reduced expression in the lymph node and CNS [417].

Prostaglandin-endoperoxide synthase 2 (Ptgs2) [often referred to as Cox2]: Increased expression of *Ptgs2* has been found in whole brain from mice fed a KD, compared to those fed a standard diet [283]; in tumour cells, the KD has attenuated a tumour-induced increase in *Ptgs2* [283] and in the hippocampus, the KD has attenuated a kainic acid-induced increase in *Ptgs2* [414].

Other inflammatory markers: Evidence regarding expression of other genes that may also serve as inflammatory markers is conflicting. No change in hepatic expression of interleukin 1 complex (*Il1*), *Il6* and chemokine ligand 2 (*Ccl2*) was found [405], whilst others reported decreased expression of *Il1b*, *Il6* and *Ccl2* in the lymph node and CNS in KD-fed mice, compared to those fed a standard diet [417].

Considering the antioxidant effects of KD-feeding observed in several studies [183], expression of genes encoding proinflammatory cytokines may be expected to be reduced with the KD – for example, *TNFA*, *IL1* and *IL6*, the related chemokine ligand 2 (expression induced by *TNFA* [426]), and *COX2*, which has been associated with excitotoxicity, ROS generation and neuronal death, [427, 428]. Seemingly contradictory findings may be explained by an initial induction of mild oxidative stress from the KD, in order to activate pathways that ultimately lead to an improved redox state (as may occur with activation of the Nrf2 pathway [412]), but this is just speculation.

Insulin-like growth factor (IGF) system: Brain IGF1 receptor, *Igf1r*, and IGF binding protein *Igfbp-3* mRNA levels were increased in rats fed a calorie-restricted KD, compared to those

fed a standard diet or a calorie-restricted diet [411]; *Igfr*, *Igfbp-2* and *-5* mRNA levels were unchanged. Others have reported increased expression of *Igfbp1* with the KD [290].

The significance of these changes for epilepsy is unknown, although the improved insulin sensitivity associated with calorie restriction (through increased adiponectin and decreased TNF α) has been proposed as one of the mechanisms by which calorie restriction results in longevity [429]. Insulin/IGF1 are also thought to play a role in Alzheimer's disease, through the regulation of APP and amyloid- β levels. [430].

Evidence regarding expression changes of mitochondrial DNA (mtDNA) is limited. The KD has been shown to increase mitochondrial biogenesis [291, 293] and upregulate transcripts encoding mitochondrial proteins [291]. The KD has also been associated with increased mtDNA copy number in skeletal muscle of wild type mice [293], as has a high-fat diet [431]. It must be noted, however, that ketosis has differing effects on skeletal muscle compared to brain [291, 432].

This review of the literature has identified a plethora of genes that are differentially expressed in KD animal models, compared to controls. When evaluating the evidence, many factors must be considered: not all findings are consistent, not all animal models used are identical and the location of gene expression changes may be significant.

No studies have been identified that have looked at gene expression changes in humans following a KD.

1.4 Research questions and hypotheses

This literature review has highlighted a need to improve targeting of the KD. KD treatment requires input from a neurologist, specialist dietitian and often a specialised nurse, it requires some level of dietary restriction (this varies according to the KD type) and can

cause adverse side effects. The ability to predict response to the KD would prevent unnecessary dietary restriction in those who are unlikely to respond and prioritise those who are more likely to respond – the KD may be used earlier in their natural history. In order to do this, predictors of response must be identified and knowledge regarding the mechanisms behind the antiepileptic/neuroprotective properties of the KD expanded. Aside from the presence of certain metabolic disorders, there is conflicting evidence regarding the influence of demographic, clinical and biochemical factors on KD response; the only really consistent finding is that neither gender nor intellectual status affect KD response. As such, in the majority of cases (those without GLUT1 or PDD), there is no mechanistic guide to KD use.

This project will investigate the genetic basis behind differential response to the KD in humans. The rationale for this is based on several factors, as outlined earlier in this chapter:

- i) a genetic basis to response to the KD for epilepsy has been shown in animal models and for Alzheimer's disease in humans
- ii) there is evidence that certain epilepsy syndromes and conditions in humans respond particularly well to the KD, many of which are caused predominantly by single genetic factors
- iii) the KD causes changes in gene expression, even if the consequences of these changes are largely unknown. Genetic variation has a major influence on gene expression and so it seems appropriate that individual genetic variation determines whether somebody responds favourably to the KD or not.

This project has two principal research questions:

Research question 1

Does common genetic variation affect response to the KD?

Research question 2

Does rare genetic variation affect response to the KD?

Work was completed in preparation for investigation of another research question

(*Research question 3*): 'what gene expression changes are induced in humans following a KD?' Due to funding limitations, this analysis could not be completed.

Project hypothesis: it is expected that there will be a differential distribution of common and/or rare gene variants, present in one or more genes, in responders of the KD compared to non-responders.

Predictors with a clinical utility (potentially a group of common variants or one or more rare variants in a gene or a group of genes) may allow dietary treatment to be considered at an earlier stage for patients who are more likely to respond favourably; dietary treatment may be avoided, or other treatment options prioritised, for those less likely to respond favourably to the KD; of course, other factors would always be considered, such as the severity of the patient's condition, treatments already trialled and the motivation of the patient and family. Even identification of genes/variants with small effect sizes on KD response would bring us one step closer to unravelling the mechanisms behind the antiepileptic effects of the KD, potentially with a view to optimising treatment in the future.

A variety of genetic analyses will be conducted, including a candidate gene analysis, genome-wide association study and whole exome sequencing. This enables the advantages of various approaches to be harnessed, optimising the chances of detecting a genotype-phenotype correlation, if one does exist.

2 Methodology

2.1 Ethics and recruitment

Ethical approval for the project was sought by the researcher and granted by the National Research Ethics Service Committee London - City Road & Hampstead (formerly East Central London Research Ethics Committee 1) (Appendix 2.1).

Based on the limited pool from which to recruit participants, an initial sample size of 100 was proposed. This was due to the small number of individuals following the KD (or who have previously followed the KD and are still under hospital follow-up) in the UK, together with the logistical difficulties in gaining ethical approval for international collaboration and ensuring consistent, continued recruitment across multiple centres/countries. The proposed sample size was later increased to 200, based on the recruitment rate. For variants of large effect size, results have been obtained from genome-wide association studies with fewer than 100 patients [433]; genetic variants that increased risk of carbamazepine-induced hypersensitivity reactions were identified from 22 patients [434]. Only a few (4-5) people [435, 436], or indeed even one individual [437], may be required for identifying causal genes with exome sequencing. Having a comparable cohort size in this project as those used in successful studies to date gives a reasonable chance of finding relevant variants, if they have a sufficiently large effect size. The power to detect a genotypic-phenotypic association, if one exists, with this sample size, depends on various factors, such as the number of genes and variants being tested and whether sequencing or genotyping was performed. More detailed power calculations are provided in each chapter. A larger cohort would increase power to detect variants with smaller effect sizes.

A favourable ethical opinion was also granted for an amendment (Appendix 2.2) to allow blood to be taken from participants prior to starting the diet (for new-starters), at one

point whilst on the diet, and when dietary treatment had been discontinued (if the participant was weaned off the diet while the study was still in operation). This would allow gene expression changes to be tracked pre-, during- and post- diet.

Research and Development approval was sought and granted for each of the following sites: Great Ormond Street Hospital for Children (GOSH), National Hospital for Neurology and Neurosurgery (NHNN), Evelina Children's Hospital (ECH), Young Epilepsy (YE), Birmingham Children's Hospital (BCH), St George's Hospital, Addenbrooke's Hospital, Alder Hey Children's Hospital, Bristol Royal Hospital for Sick Children (BRHSC) and The Royal Children's Hospital in Melbourne (RCHM). The researcher obtained honorary research contracts for each site, except for RCHM, where Jacinta McMahon provided phenotypic data.

The researcher attended every KD clinic (both new patient and follow-up clinics) at GOSH, ECH and NHNN during the recruitment period. Patients who had previously followed the KD at these hospitals were also identified (from lists of past KD patients kept by dietitians, as well as from a list provided by Prof Helen Cross of participants from the GOSH RCT) and the researcher recruited from specialist clinic appointments attended by any of these individuals who were still under follow-up. The researcher sought permission from, or made the consultant aware that recruitment was taking place during his/her clinic. At several sites, potential participants were identified and recruited by specialist/research nurses (at BCH and BRHSC), a research assistant (at RCHM), keto-assistants (at Matthew's Friends clinics, held at Young Epilepsy) or dietitians (at St George's).

Inclusion criteria:

- Individuals aged ≥ 3 months who were either following the KD, who were soon to be commencing the KD, or who had followed the KD in the past for his/her epilepsy.

Exclusion criteria:

- Individuals who discontinued the diet before the 3-month point due to lack of tolerability (those who discontinued the diet before the 3-month point due to lack of response or seizure increase were not excluded).
- Individuals with known GLUT1 deficiency, PDD or other metabolic disorders.
- Individuals with progressive myoclonic epilepsies (as lack of response may be due to the progressive nature of the condition).

Potential participants were approached by the researcher or other recruiter, either in the waiting room, prior to or following the clinic appointment, or during the actual clinic appointment. Parents/guardians/patients were politely asked if they would mind sparing a few moments and were offered to go to a quiet area (such as a free clinic room, or the nearby playroom) to discuss the study. Adult patients, parents/guardians, young people and children were provided with an age-specific, written information sheet, adapted to whichever facility in which the patient was receiving care, explaining the study and all procedures involved. It was emphasised that participation was entirely voluntary. The researcher or the patient's consultant took consent; parental consent was obtained for patients under 16 years of age and assent forms were also provided for children. Sample information sheets and consent/assent forms for adult patients at the NHNN and for parents/guardians, young people and children at GOSH are provided in Appendix 2.3. Everybody had the opportunity to discuss any issues pertaining to their/their child's participation in the study before signing the consent form. Consent was offered within the clinic, but people could wait until the subsequent clinic visit. The use of interpreters was made available where required. Participants received no compensation from having taken part in the study.

A breakdown of the number of participants obtained from each site, and whether they were recruited prospectively or retrospectively, is given below:

- GOSH: n=93 of which 58 were prospective and 35 were retrospective. 77 were eligible for inclusion in the study.
- NHNN: n=18 of which 15 were prospective and 3 were retrospective. 14 were eligible for inclusion in the study.
- ECH: n=82 of which 57 were prospective and 25 were retrospective. 67 were eligible for inclusion in the study.
- YE in conjunction with Matthew's Friends: n=13 of which 8 were prospective and 5 were retrospective. 11 were eligible for inclusion in the study.
- BCH: n=24 of which 23 were prospective and 1 was retrospective. 21 were eligible for inclusion in the study.
- St George's Hospital: n=11, all of which were prospective. 8 were eligible for inclusion in the study.
- Addenbrooke's Hospital: n=19 of which 7 were prospective and 12 were retrospective. 19 were eligible for inclusion in the study.
- Alder Hey Children's Hospital: n=4, all of which were prospective. 4 were eligible for inclusion in the study.
- BRHSC: n=2, all of which were prospective. 1 was eligible for inclusion in the study.
- RCHM: n=35 of which 6 were prospective and 29 were retrospective. 31 were eligible for inclusion in the study.

The researcher sat in every clinic for all patients at GOSH, ECH and NHNN who were recruited prospectively; with all retrospectively-recruited patients from GOSH, ECH and NHNN, the researcher discussed their previous experiences with the KD in a quiet place either before or after clinic. The researcher attended clinic for 4 participants from Matthew's Friends, and contacted 4 participants by postal letter; the researcher attended clinic for 6 of the participants from BCH; the researcher attended two clinics at St George's Hospital. The researcher attended no clinics at BRHSC, RCHM or Alder Hey Children's Hospital. Dr Alasdair Parker and Dr Anna Maw from Addenbrooke's Hospital, with assistance from the dietitian Helena Champion, provided the researcher with a list of contact details of patients who had previously followed the KD – the researcher was given

permission to contact the parent/guardian(s) of these patients to ask whether they wished to participate in the study. Following an initial telephone conversation, patient information sheets and consent forms were sent to parent/guardian(s) and, if they wished to participate, DNA stored at Addenbrooke's was used for genetic analyses.

Patients following the KD at the time of recruitment were asked to complete a questionnaire regarding compliance to the KD (patient and parent/guardian versions - used for patients unable to complete the questionnaire - are given in Appendix 2.4). This questionnaire was adapted from a validated questionnaire used to assess compliance with AEDs at The Epilepsy Society's Chalfont Centre, Chalfont St Peter. It is based on the Medication Adherence Report Scale (MARS) developed by Professor Rob Horne from UCL School of Pharmacy. The MARS questionnaire was designed for measuring adherence to medicines used for psychiatric disorders, but it has been widely used for a range of conditions. It measures non-adherence using a five-item Likert scale, which accounts for the fact that non-compliant behaviour extends across a spectrum [438]. The questionnaire also attempts to address non-adherence in a non-threatening and non-judgemental manner, encouraging patients to answer truthfully [439]. The KD version of a questionnaire was registered as an audit at GOSH, NHNN, YE, BCH, Addenbrooke's Hospital and St George's Hospital, and as a service evaluation at ECH.

2.2 Phenotypic data collection

Phenotypic data collection was obtained from patient hospital records (paper and electronic versions, including main hospital records, Dietetic department records and Neurology department records), using a proforma created by the researcher (Figure 2.1).

Unique identification code		
Clinical history (including pregnancy/birth details, onset of seizures, history of head injury, history of status epilepticus)		
Height	Pre-diet:	3 month point: 6 month point: 12 month point: 18 month point: 24 month point:
Weight	Pre-diet:	3 month point: 6 month point: 12 month point: 18 month point: 24 month point:
Epilepsy syndrome		
Seizure types (before and during KD)		
AEDs/drugs	Pre-diet:	3 month point: 6 month point: 12 month point: 18 month point: 24 month point: Previous (start date, stop date, max dose, response and side effects):
Family history (epilepsy and other)		
Neurodevelopment (including formal developmental assessments)		
Neurological findings		
EEG data		
Imaging data		
Date started KD and diet type		

Date stopped KD (if retrospective)		
Response to KD	3 month point: 6 month point: 12 month point: 18 month point: 24 month point:	
Biochemical data	Pre-diet:	3 month point: 6 month point: 12 month point: 18 month point: 24 month point:
Side effects		
Long-term outcome	Post-2 years:	

Figure 2.1: Phenotype proforma

All data were pseudo-anonymised prior to entry into a database, with a unique identification number allocated and saved elsewhere. Only those directly involved in the study had access to information. All personal data were stored on a secure NHS computer and were password-protected. Hard copies were kept on NHS premises in a locked cabinet.

2.3 Genotypic data collection

2.3.1 Research questions 1 and 2

(Does common or rare genetic variation affect response to the KD?)

For participants who were following the KD or who had already weaned off the diet: participants had blood taken once, usually at the same time as blood taken for routine clinical monitoring. Anaesthetic cream or spray was used where appropriate and blood was taken by an experienced phlebotomist. 2-5ml was taken from children, 10ml from adults, and put into an EDTA (ethylenediaminetetraacetic acid) tube.

EDTA samples were sent to the Molecular Genetics laboratory at GOSH, where DNA was extracted by clinical geneticists.

DNA was extracted by one of two methods, depending on workflow and the volume of blood obtained: for blood samples 1.5ml-4ml, Autogen (AutoGen Inc, Hollister, Massachusetts, USA) was used; for blood samples 0.5ml-2.5ml, Fujifilm (FUJIFILM corporation, Tokyo, Japan) was used. Autogen is an automated machine which uses Qiagen's FlexiGene procedure – a protease digestion/DNA precipitation method, whereby the erythrocytes and leucocytes are lysed with a buffer while keeping the nuclei intact. A nuclei pellet is obtained by centrifugation and DNA precipitated by addition of isopropanol. Fujifilm uses pressured filtration technology, which is carried out using an 80µm thick porous membrane that immobilises the nucleic acid without need for centrifugation.

DNA samples were also received from RCHM, where DNA had been extracted from blood using Maxi Qiagen kits, which use the FlexiGene procedure.

All DNA samples were vertically stored in a -80°C freezer at UCL Institute of Neurology.

DNA concentration and 260/280 ratios (a measure of DNA purity - for pure DNA, the ratio of the absorbance at 260 and 280nm is approximately 1.8 [TECHNICAL SUPPORT BULLETIN, NanoDrop® ND-1000 and ND-8000 8-Sample Spectrophotometers, www.nanodrops.com]) of all samples was measured with NanoDrop ND-1000 Spectrophotometer (Thermo Fisher Scientific Inc., USA).

At each of the study sites, it is standard practice to test children with infantile-onset epileptic encephalopathy for GLUT1 deficiency (particularly if they have a very favourable response to the KD), either from *SLC2A1* mutation, hypoglycorrachia or low cerebrospinal fluid:blood glucose ratio. In case any individuals had not been tested, *SLC2A1* was sequenced in all samples to be included in genetic analyses at the North East Thames Regional Genetics Laboratory by Dr Suzanne Drury to formally exclude the possibility of GLUT1 deficiency. DNA was quantified using a Qubit® Fluorometer (Life Technologies

Corporation). 250ng of DNA was used in the Illumina TruSeq Custom Amplicon Library Preparation protocol, according to manufacturer's instructions (TruSeq Custom Amplicon Library Preparation Guide, © 2013 Illumina, Inc). This includes hybridisation of oligo probes (designed to target the 10 coding exons of *SLC2A1*), removal of unbound oligos, extension-ligation of bound oligos to form DNA templates that consist of the regions of interest flanked by primer sequences, and finally polymerase chain reaction (PCR) amplification of the DNA templates, incorporating indexes to enable sample pooling. Libraries were normalised in order to ensure equal representation of each sample. Up to 96 samples were then pooled, prior to sequencing on an Illumina MiSeq® sequencing system. This uses 'sequencing by synthesis' technology, where four fluorescently-labelled nucleotides are added in each sequencing cycle, one of which attaches to the chain of DNA; the emitted fluorescence (identified from laser excitation) identifies each base; the data are then aligned and compared to a reference genome. For each sample, paired-end sequencing reads were mapped to the February 2009 assembly of the human genome (Genome Reference Consortium Human Build 37) using the Burrows-Wheeler Aligner (BWA) and the default settings [440]. Sequence variants were called using VarScan2 (mpileup2cns) with a minimum total read depth requirement of 30, and the default settings [441]. Variants were limited to coding regions and 14 and 6 base pairs 5' and 3' respectively. Variants were annotated using the Ensembl Variant Effect Predictor Perl Script.

No individual with unknown GLUT1 deficiency harboured a missense variant that was predicted to be damaging. Two individuals (one extreme non-responder who discontinued the KD immediately and one with a variable response to the KD, who remained on the KD long-term) harboured a missense variant in *SLC2A1* but these were both predicted to be tolerated by functional prediction algorithms. All synonymous and non-coding variants with a minor allele frequency (MAF) <2% were analysed with Alamut (Interactive Biosoftware,

LLC), but none were predicted to affect splicing (removal of intronic regions located between exons for production of RNA).

SLC2A1 sequencing failed in eight individuals due to low quantity or quality DNA.

Samples were also prepared by the researcher to be sent to

- a) AROS Applied Biotechnology A/S, Denmark to be genotyped with the *Infinium HumanOmniExpressExome Beadchip* (Illumina Inc, San Diego, USA) (all cases with DNA samples eligible for inclusion as of January 2013)
- b) Royal Devon & Exeter Hospital for sequencing of *KCNJ11* and *BAD* (all cases with DNA samples eligible for inclusion as of April 2013)
- c) Institute of Neurology Neurogenetics Unit for whole exome sequencing (45 selected extreme responders and non-responders)

Further information regarding Methodology, including sample selection and genetic analyses, will be described in Chapters 4-7.

2.3.2 Research question 3

(What gene expression changes are induced in humans following a KD?)

For new starters: participants had blood taken at one point pre-diet and at one point when following the diet (preferably at the 3-month point). Samples were usually taken at the same time as blood taken for routine clinical monitoring and the same procedure was used as for Research questions 1 and 2. Although ethical approval was also obtained for post-diet blood samples, the logistical difficulties of this quickly became clear – patients often weaned off the diet before their next clinic appointment.

On each blood-letting occasion, 2.5ml blood was put into a PAXgene Blood RNA tube (PreAnalytix GmbH, Switzerland), for reduced RNA degradation and gene induction [442].

On one of these occasions, an extra (approximate) 2.5ml blood was put into an EDTA tube for DNA extraction, as described above.

PAXgene tubes were softly inverted 8-10 times to mix the sample with the preserving agent, kept at room temperature for 2-6 hours (in order to ensure complete lysis), then at 4°C overnight. The following morning, tubes were put in a -80°C freezer at UCL Institute of Neurology until processing.

Prior to RNA extraction, PaxGene tubes were left to defrost at room temperature overnight (approximately 14 hours). Total RNA was manually extracted by Dr Mar Matarin, following the PaxGene Blood RNA Kit protocol (Qiagen, USA; reference number: 762174). Samples were processed in two batches, both including an equal number of pre-diet and on-diet samples from randomly selected patients. Nucleic acid pellets in the PAXgene tube were isolated by centrifugation, washed, re-suspended by vortexing, and incubated with Proteinase K and lysis Buffer BR2 to bring about protein digestion. Following transferral of the supernatant to a fresh microcentrifuge tube and an additional centrifugation to remove residual cell debris, ethanol was added and the tube was loaded onto a PAXgene RNA spin column. When centrifuged, RNA selectively binds to the PAXgene silica-gel membrane and contaminants pass through. After removing contaminants with several wash spins, elution with Buffer BR5 pipetted directly onto the PAXgene column membrane was undertaken, followed by incubation. RNA was chilled on ice, ready for storage. RNA concentration and 260/280 ratios (a measure of RNA purity - for pure RNA, the ratio of the absorbance at 260 and 280nm is >1.8-2.0 [443]) was measured using NanoDrop ND-1000 Spectrophotometer (Thermo Fisher Scientific Inc., USA).

RNA samples (n=42, one pre-diet and one on-diet for 21 individuals) are currently in storage at -80°C at UCL Institute of Neurology. Pending funding, it is planned to analyse

these samples with the Affymetrix Human Exon 1.0 Sense Target Array (or a similar array).

This will allow investigation of gene expression levels and alternative splicing in individuals following the KD, compared to pre-diet levels, and to determine whether these differ in responders and non-responders.

3 Phenotypic results

3.1 Introduction

A phenotype is defined as the ‘observable traits of the organism’ [444], or often, in medical contexts, as ‘deviation from normal morphology, physiology, or behaviour’ [445].

Establishing the phenotype of a patient is a challenging, yet critical task, both for clinical management and observational and interventional research. The ability to detect an association of a phenotypic trait with another factor, and the confidence in the validity of this association, depends on the accurate delineation of the trait in question on a case-by-case basis [446]; incorrect classification of cases/controls or, in this case, responders/non-responders, may result in false negative or false positive findings. To this purpose, the definition of the phenotypic trait in question should be carefully considered and high quality standards should be adhered to in order to produce accurate and reproducible results, especially if inter-study comparison is desired at a later stage. The very nature of phenotypic information, which may change over time and can be subject to external influence, as well as the methods used to obtain this information, may hinder this accuracy. This highlights the importance of extending phenotypic measurements over time and including information regarding environmental factors and other biological systems aside from those that are obviously involved with the phenotypic trait in question [447]. A balance must be sought between what is thought to be relevant (although this is debatable, for example, in the context of the KD, where evidence regarding factors that affect response is conflicting), completeness of the data sources, and time available for phenotyping.

Deep phenotyping, defined as the ‘precise and comprehensive analysis of phenotypic abnormalities in which the individual components of the phenotype are observed and

described' [445], reflects the understanding that disease, or phenotypic traits, often represent a spectrum. Even if the principal outcome measured in a study is binary (affected/unaffected, or responder/non-responder), other factors may be involved. In the context of the KD, the effects on diet response of various factors, such as age at seizure onset, could be assessed and relevant covariates could be included in analyses. This could not be done without detailed documentation of demographic and clinical data for each participant.

Deep phenotyping is especially important in genetic studies, where the functionality of variants with a genotypic-phenotypic association may be unknown. The extent of the effects of a particular genotype on a particular trait may be better understood; for example, whether homozygosity at a particular locus has a more mild or severe effect on the phenotype, compared to heterozygosity. Such detailed understanding may be especially challenging, but nevertheless important, with rare variants or those with small effect sizes. Elucidating influences on the phenotype may not necessarily lead to results with clinical utility, but will certainly further understanding of the network of genotypic-phenotypic associations, including gene-gene or pathway interactions.

Here, an overview of the cohort is presented. Information already obtained as part of the KD service provision, along with questionnaires handed out to participants recruited prospectively, were used to classify KD response, obtain rough measures of compliance and detailed clinical and demographic data. KD response was classified in terms of seizure frequency, both at specific time points and as a summary of longitudinal response; other potential contributors to what constitutes a clinically meaningful response were also considered. Potential correlations of clinical or demographic parameters with KD response were assessed. This allows comparison of response rates and demographics with other

published cohorts and encourages thought as to how phenotypic information can be used for genetic analyses.

3.2 Methods

As outlined in Chapter 2, clinical and demographic data were obtained from hospital notes, using a proforma created by the researcher.

Where possible, it is important to assess compliance to the KD to ensure that cases are not classified as non-responders simply because they did not adhere to treatment in the first place, rather than due to a particular biological/genetic reason. The more patients with inadequate compliance that are included in a clinical trial, the lower the power [448], and so a larger sample size would be needed to obtain the same possibility of committing Type I (false positive) or II (false negative) error. This notion can feasibly be extended to association studies. In this study, compliance to the diet was measured by serum BHB (this is done anyway as part of standard KD monitoring and measurements were subsequently obtained from hospital records) and questionnaires, adapted from a validated questionnaire previously used to assess compliance with AEDs (a copy is given in Appendix 2.4). Answers to questions 1-6, solely from patients who received the diet orally, were used to create a compliance score for each case: if the answer to a question was 'never', a score of 5 was given; 'rarely' = 4; 'sometimes' = 3; 'often' = 2; 'very often' = 1. The maximum compliance score was 30. Cronbach's alpha, a measure of internal consistency for questionnaires (to determine whether questions that propose to measure a similar construct produce similar scores) expressed as a number between 0 and 1, was calculated using the 'cronalpha' function in Microsoft Excel 2010 (v. 14, Microsoft, Washington, USA). Urine and/or blood ketone levels (which are often checked on a daily basis by parents, particularly during the first few months of dietary treatment) would have been useful to

assess compliance, but data were seldom present in hospital records or dietetics notes; where data were available, measurements were often sporadic and not documented over a prolonged period of time.

KD response was defined in terms of seizure-frequency. Seizure frequency was calculated in four week (28 day) epochs prior to starting the diet (baseline), prior to the 3-month point, 6-month, 12-month, 18-month and 24-month point, where applicable. These time points represent when follow-up with the consultant neurologist usually occurs after starting the KD; weaning of the diet is typically considered after two years. If clinic appointments fell >2 weeks before or after a specific follow-up point, diet response at that point was classified as unknown. Where seizure diaries were completed as part of monitoring for the KD service provision, these were used to calculate exact seizure frequency. Otherwise, clinic letters prior to starting the diet and at each KD follow-up clinic were used to gain a crude measure. Response to the diet after 24 months was noted for participants with long-term follow-up, solely from clinic letters.

Cases were placed into a response category at each follow-up point. Cases with $\geq 50\%$ seizure reduction were classified as 'responders'; those with $< 50\%$ seizure reduction were 'non-responders'. A $\geq 50\%$ seizure reduction is viewed as clinically useful in this drug-resistant cohort (personal communication with GOSH Paediatric Neurologist). Personal contact between many patients/parents/guardians and the researcher through attendance at clinic appointments aided diet response classification – the researcher witnessed, first-hand, any visible changes in patients since starting the diet and heard detailed descriptions from parents/guardians regarding their experiences with the diet. Such details may not always be mentioned in clinic letters.

A Chi-square goodness of fit test was used to assess whether diet response at a single time point correlates with response at other time points. This tests the null hypothesis that the observed number of matches (responder-responder, or non-responder-non-responder) is different to the expected number of matches at two particular time points, given the proportions of responders/non-responders at these time points, and an assumption of independence. A significance threshold of 0.05 represents the probability that the observed distribution of data is due to chance. With a p-value of <0.05 , the null hypothesis of independence (the observed number of matches is different to the expected number of matches) can be rejected, with a less than 1 in 20 chance of being incorrect.

The merits of classifying diet response for each case at specific follow-up points, versus placing each case into a summary response category that summarised his/her diet response over time, were considered. Two summary response categories were created:

- i) Summary diet response, where cases are classified as responders or non-responders, based on their response at all time points recorded
- ii) Extreme summary diet response, which aims to include those cases with the most extreme response to dietary therapy.

Logistic regression was used to assess the effect of various demographic and clinical factors on diet response, based on the summary diet response phenotypes. Analyses were re-conducted, using diet response at specific follow-up time points, for those factors with a statistically significant correlation with summary diet response. For categorical factors, pairwise combinations not reported from the original calculation (one category is automatically used as a base category) were analysed using the 'relevel' function, which redefines the base category. For example, in the original calculation comparing diet response and diet type, the classical KD was automatically defined as the base category and

so response in cases following the classical KD was compared to those following either the MCT KD or the MAD. By redefining the base category as either MAD or MCT KD, response in cases following these two diet types could be compared. All analyses assessing putative correlations between diet response and demographic/clinical factors will help indicate what variables, if any, to include in genetic analyses.

To avoid the inflated likelihood of Type I error when testing multiple hypotheses, a Bonferroni-corrected significance threshold was calculated, based on an alpha of 0.05 and the number of tests conducted. This is a simple and widely-used method of correcting for multiple testing in statistical analyses, although (for non-genetic analyses) there is no accepted gold standard [449].

Clinic letters and dietetics notes were used to extract information regarding diet response beyond seizure frequency, but these beneficial/adverse effects were not quantified.

Microsoft Excel 2010 (v. 14, Microsoft, Washington, USA) and R: A Language and Environment for Statistical Computing (v. 3.0.0, R Foundation for Statistical Computing, Vienna, Austria) were used for descriptive and inferential statistics.

3.3 Results

3.3.1 An overview

Figure 3.1 shows the number of cases recruited and included in final phenotypic analyses.

This includes all cases, independent of quality control criteria employed in genetic analyses.

The cohort clinical and demographic data are summarised in Table 3.1.

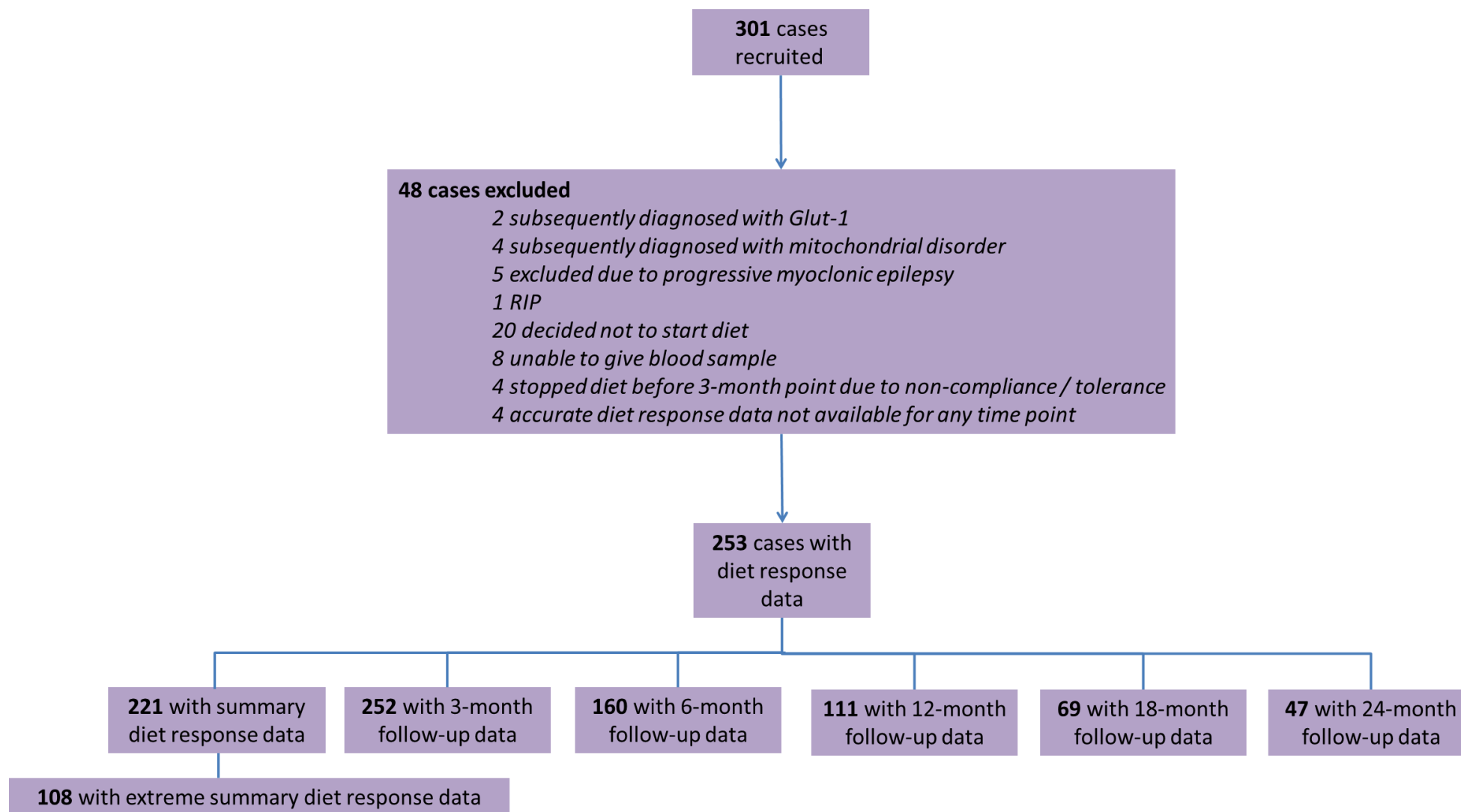


Figure 3.1: Flowchart of cases recruited and included in final analysis

Table 3.1: Cohort clinical and demographic data (for cases with diet response data, n=253)

Gender	Male n=132 (52%)	
	Female n=121 (48%)	
Age at seizure onset (years) median(IQR)	0.67(1.86) (unknown for 1 case)	
Age at diet onset (years) median(IQR)	5.60(6.60)	
‘Developmental delay’ or ‘learning difficulties’ (as described in clinic letters)	Cases described as having developmental delay n=168 (66%)	
	Cases described as having learning difficulties n=37 (15%)	
	No developmental delay/learning difficulties n= 18(7%)	
	Unknown n=30 (12%)	
Cause of epilepsy *	Genetic n=29 (11%)	Mutation in <i>SCN1A</i> n=14
		Mutation in <i>CDKL5</i> n=3
		Mutation in <i>PCDH19</i> n=1
		Mutation in <i>MECP2</i> n=3
		Mutation in <i>STXBP1</i> n=1
		Chromosome translocation n=2
		17q22–q23.1 deletion n=1
		17p13.3 deletion (Miller-Dieker syndrome) n=1
		3p25 deletion n=1
		Chromosome q22 deletion n=1
		Cardio-facio cutaneous syndrome (unknown which gene) n=1
	Structural-metabolic n=72 (28%)	Hypoxic ischemic encephalopathy (HIE) n=20
		Perinatal stroke (without HIE) n=1
		Haemorrhage n=4
		Neonatal hypoglycaemia n=2
		Polymicrogyria n=10
		Pachygyria n=1
		Infection n=6
		Periventricular leukomalacia n=1
		Cystic encephalomalacia n=1

	Occipital lobe abnormality n=1
	Temporal lobe abnormality n=3
	Hypoplasia n=2 (1 with cerebellar hypoplasia and 1 with hypoplasia of the corpus callosum)
	Cortical dysplasia n=2
	Porencephaly n=1
	'Lesions' n=1
	Neurofibromatosis type 1 n=1
	Epidermal naevus syndrome n=1
	Tuberous sclerosis n=5
	Hippocampal sclerosis n=2
	Lissencephaly n=2
	Ohtahara syndrome n=1
	Unknown n=4
Unknown cause n=152 (60%)	
Epilepsy syndrome *	Dravet syndrome/severe myoclonic epilepsy of infancy n=15 (6%)
	Lennox-Gastaut syndrome/LGS-spectrum n=13 (5%)
	Childhood absence epilepsy n=3 (1%)
	Juvenile myoclonic epilepsy n=2 (1%)
	Juvenile absence epilepsy n=3 (1%)
	Epilepsy with myoclonic atonic (astatic) seizures (Doose syndrome) n=14 (6%)
	Epilepsy with myoclonic absences n=1 (1%)
	Epilepsy with myoclonic atonic seizures and myoclonic absences n=1 (1%)
	Myoclonic epilepsy (unspecified) n=7 (3%)
	Epilepsy of infancy with migrating focal seizures n=3 (1%)
	Ohtahara syndrome n=1 (1%)
	West syndrome n=15 (6%)
	Undiagnosed n=175 (69%)
Seizure localisation based on EEG and clinical features	Generalised n=78 (31%)
	Focal n=91 (36%)
	Focal with secondary generalisation n=7 (3%)
	Focal and generalised (cases with seizures that originate from a specific area in one hemisphere, as well as seizures that originate from bilateral hemispheres, as shown by EEG) n=71 (28%)

	Spasms n=5 (2.0%)
	Unknown = 1 (0.4%)
Number of AEDs at diet onset <i>mean[95% CI]</i>	2.34[2.22, 2.46]
AEDs at diet onset **	Sodium valproate n=131 (52%)
	Levetiracetam n=81 (32%)
	Topiramate n=68 (27%)
	Vigabatrin n=23 (9%)
	Carbamazepine n=21 (8%)
	Lamotrigine n=56 (22%)
	Clobazam n=95 (38%)
Number of failed AEDs prior to diet onset <i>mean[95% CI]</i>	6.61[6.28, 6.94] (unknown for 3 cases)
Diet type (at 3-month point) ***	Classical Ketogenic diet n=166 (66%)
	Medium Chain Triglyceride Ketogenic Diet n=47 (19%)
	Modified Atkins Diet n=39 (15%)
	Unknown n=1 (0.4%)
Feed	Oral n=173 (68%)
	Gastrostomy/naso-gastric tube n=64 (25%)
	Oral and tube n=16 (6%)

* Cause of epilepsy (*genetic, structural/metabolic, unknown*) and epilepsy syndromes have been classified according to Berg et al. (2010)

** Only the most common AEDs taken in our cohort have been included in 'AEDs at diet onset'; other AEDs (such as nitrazepam or gabapentin) were only taken by a small number of cases, which did not allow for stratification into responders and non-responders - there would be too few cases in each group for statistical analysis.

*** no patients were following the Low Glycaemic Index Treatment, as this was not offered as a diet option at the study sites. If a patient transitioned to a different diet type before the 3-month point, the new/second diet type was considered this individual's 'diet type'.

3.3.2 Compliance

Compliance questionnaires were returned from 66 patients, either from patients themselves (five cases) or from parent/guardian(s). The mean [95% confidence interval] compliance score out of 30 was 26.5 [25.7, 27.3]; follow-up time ranged from 2-82 months. There is no standard cut-off point for high/low compliance from such questionnaires; the threshold is based on the frequency distribution for the particular dataset. As the distribution of compliance scores is positively skewed (shown in Figure 3.2) and there is no natural cut-off point where the group can be dichotomised, the lower 3rd was taken as the 'low compliance' group. Approximately one quarter of patients with compliance questionnaire data scored $\leq 25/30$ (the lowest score was 17/30).

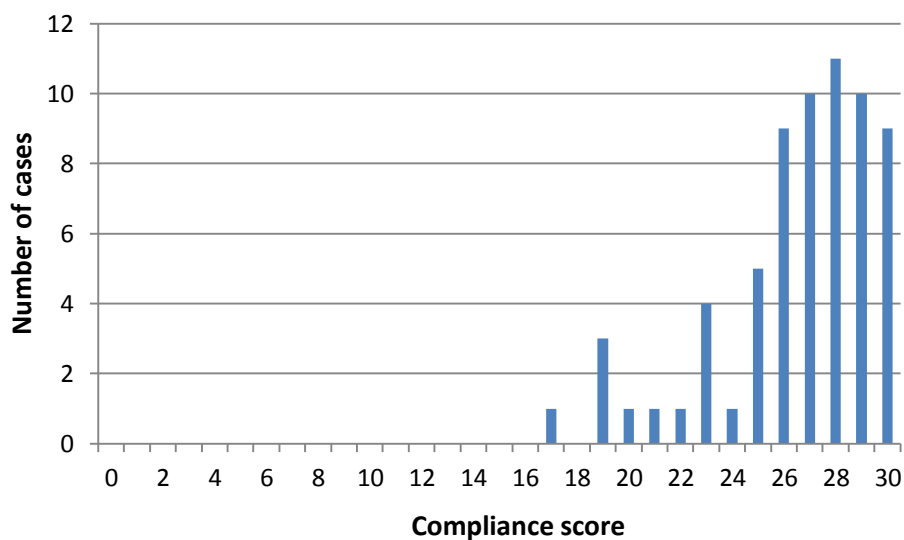


Figure 3.2: Distribution of diet compliance scores, n=66

Cronbach's alpha was 0.68. This means that the questions in the compliance questionnaire have an average intercorrelation of 0.68, or that 68% of the variance is consistent across the six questions.

A coefficient of 0.68 may be viewed as low, considering that a value of 0.8 is generally accepted as reliable. This has, however, been disputed - values of 0.7 or below may be appropriate, depending on what is being measured [450]. Cronbach's alpha depends on several factors, such as the number of items in the scale, especially when intercorrelation between items is low [451]. A low alpha value may also be attributed to heterogeneous question constructs [452] or skewed distribution of data scores [453]. The small number of items in the scale and the positively skewed distribution of compliance scores may have lowered the alpha value.

As shown in Figure 3.3, there is no clear correlation between length of time on diet and compliance levels.

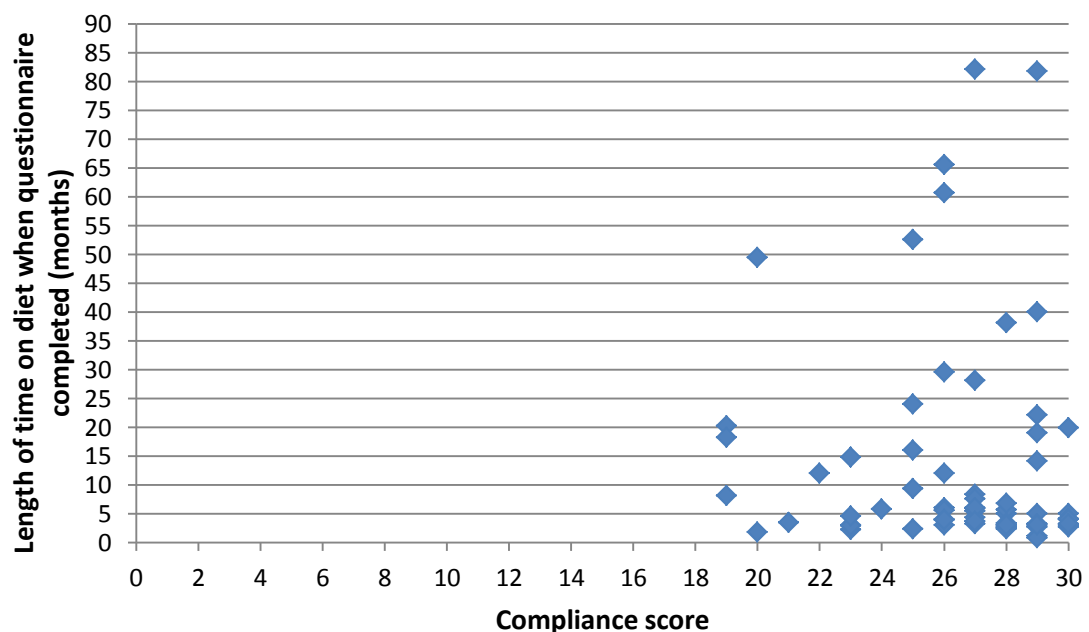


Figure 3.3: Scatter plot of compliance scores and length of time on diet when questionnaire was completed, for 66 individuals

Serum BHB levels were available for:

122 cases (48% of those with 3-month diet response data) at 3-month follow-up:

median(IQR) 2.74(2.61)mmol/L

70 cases (44% of those with 6-month diet response data) at 6-month follow-up:

median(IQR) 3.06(2.58)mmol/L

56 cases (52% of those with 12-month response data) at 12-month follow-up: *median(IQR)*

2.89(2.66)mmol/L

Data were not normally distributed at any time point.

BHB was not recorded for 18- or 24-month follow-up, due to small sample sizes.

Considering that the 'recommended' (in a handbook of Dietary Treatments for Epilepsy, but not supported by published studies) serum BHB level for individuals following the KD is 4-6mmol/L [454], the median BHB levels in our cohort appears low. However, the interquartile ranges (a measure of dispersion and variability in observations) are wide, meaning that the middle 50% of observations are widely spaced apart [455]. 41(34% of 122) cases had serum BHB levels ≥ 4 mmol/L at 3-month follow-up, 24(34% of 70) cases at 6-month follow-up and 20(36% of 56) cases at 12-month follow-up. The largest proportion of patients with BHB levels within the normal range (<0.6 mmol/L) was 9/122(7%), at 3-month follow-up. Potential reasons for variability in serum BHB levels are outlined in the Discussion.

3.3.3 Ketogenic diet response: seizure frequency at specific follow-up points

The proportions of responders/non-responders at 3-month, 6-month, 12-month, 18-month and 24-month follow-up are given in Figure 3.4, Figure 3.5, Figure 3.6, Figure 3.7 and Figure 3.8 respectively. A proportion of participants were unclassifiable at specific follow-up points (range 2-9%), either due to insufficient information provided in clinic letters or missed appointments.

9(4% of 252) participants were seizure-free at the 3-month point, 10(6% of 160) at 6-months, 7(6% of 111) at 12-months, 3(4% of 69) at 18-months and 3(7% of 46) at 24-months.

Diet response was calculated from seizure diaries in 35/252(14%) participants at the 3-month point, 23/160(14%) at the 6-month point and 14/111(13%) at the 12-month point.

No participants had completed seizure diaries at the 18- and 24-month points.

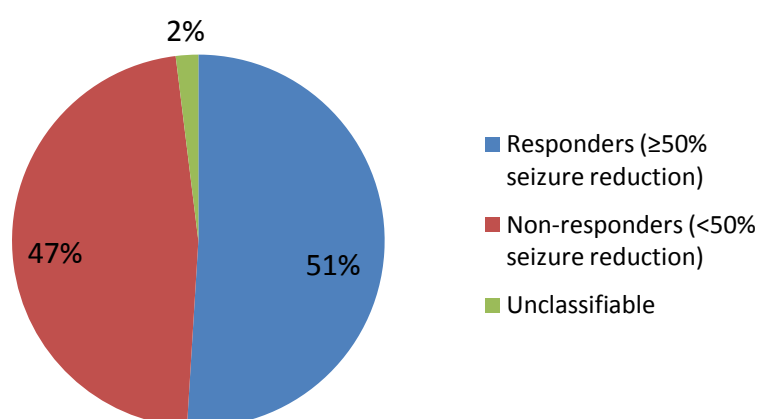


Figure 3.4: 3-month diet response, n=257 (n=252 with classifiable diet response data)

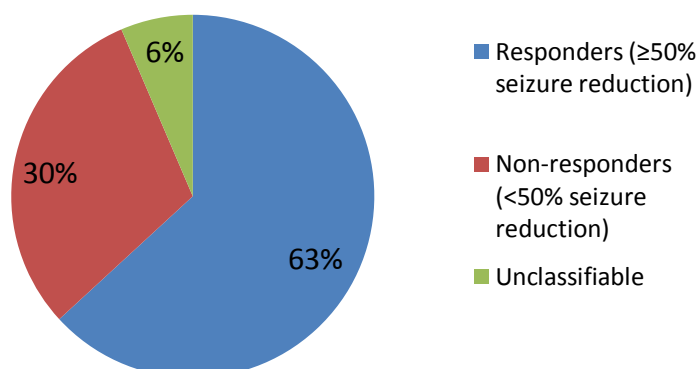


Figure 3.5: 6-month diet response, n=171 (n=160 with classifiable diet response data)

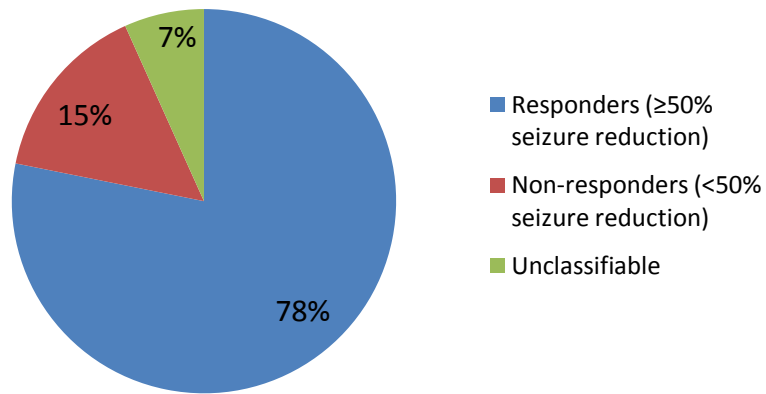


Figure 3.6: 12-month diet response, n=119 (n=111 with classifiable diet response data)

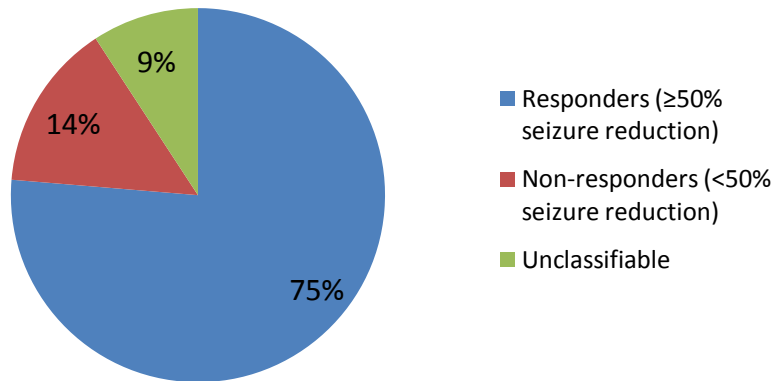


Figure 3.7: 18-month diet response, n=77 (n=69 with classifiable diet response data)

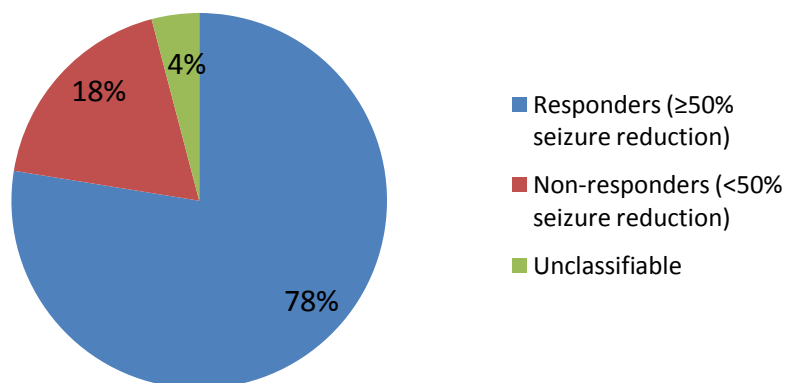


Figure 3.8: 24-month diet response, n=49 (n=46 with classifiable diet response data)

If an intent-to-treat type approach were adopted (to the extent that is possible in this study, as not all people who were referred for the KD at each centre were identified and approached), thus including individuals who were recruited but then decided not to start the KD (n=20), those who stopped the diet before the 3-month point due to non-compliance or adverse side effects (n=4), or who died whilst following the KD (n=1), response rates are as follows:

3-month point: 131/282(46%) achieved $\geq 50\%$ seizure reduction, of which 9(3%) were seizure-free.

6-month point: 108/282(38%) achieved $\geq 50\%$ seizure reduction, of which 10(4%) were seizure-free.

12-month point: 93/282(33%) achieved $\geq 50\%$ seizure reduction, of which 7 (2%) were seizure-free.

18-month point: 58/282(21%) achieved $\geq 50\%$ seizure reduction, of which 3(1%) were seizure-free.

24-month point: 38/282(13%) achieved $\geq 50\%$ seizure reduction, of which 3(1%) were seizure-free.

As may be expected, most cases who continue to follow the KD long-term are classified as responders. Some cases, however, defined as non-responders in terms of seizure frequency, remained on the diet, even up to 24 months after starting. Potential reasons for this are outlined in the following section.

3.3.4 Factors beyond seizure frequency

Other benefits on seizures, aside from reduced frequency, reported at any time point, are given in Table 3.2.

Table 3.2: Reported benefits on seizures, aside from reduced seizure frequency, from the Ketogenic diet

Benefit	Number of cases who reported benefit, n (% of 253[cases with diet response data])
Less severe seizures (defined as less 'severe' or 'milder' in hospital records)	n=36 (14%)
Shorter seizures	n=26 (10%)
Reduced used of emergency medication	n=17 (7%)
Quicker recovery time	n=15 (6%)
Shorter clusters of seizures	n=11 (4%)
Fewer daytime seizures	n=4 (2%)
Fewer hospital admissions	n=4 (2%)
Fewer episodes of status	n=4 (2%)
Fewer convulsive seizures/seizures that lead to falls	n=3 (1%)
Fewer night-time seizures	n=2 (1%)
Now able to recognise warning signs before a seizure	n=2(1%)
Seizures don't 'progress' to the same extent as before	n=2 (1%)
Seizure pattern changed to clusters	n=1 (0.4%)

40 individuals who reported one or more of these benefits were classified as non-responders at the 3-month point. This represents 33% of the total number of non-responders at this time point. 23/40(57.5% of the 40 who reported one or more seizure-related benefits) cases remained on the diet at the 6-month point; 15/40(37.5%) cases remained on the diet at the 12-month point; 11/40(28%) cases remained on the diet at the 18-month point; 5/40(12.5%) cases remained on the diet at the 24-month point.

For those who followed a KD for six months or longer, 52/164(32%) patients were able to reduce their number or dose of AEDs without an adverse effect on seizures (AEDs are generally not changed before the 3-month point); seven of these 52 patients were weaned off all regular AEDs - this was still the case at the most recent follow-up (range: 1-9 years

since AED-free, whilst still following the KD), with the exception of one case who recently had an AED introduced due to worsening seizure control, with no effect (AED to be weaned). Four further patients are currently reducing AED dose, the effect of which is still unknown.

Other non-seizure-related benefits reported at any time point, are given in Table 3.3. 38 cases (31% of the total number of non-responders at the 3-month point) who reported one or more of these benefits were classified as non-responders at the 3-month point. 22(58% of 38) of these cases remained on the diet at the 6-month point (five were classified as responders at this point); 14(37% of 38) remained on the diet at the 12-month point (eight were classified as responders at this point); 10(26% of 38) remained on the diet at the 18-month point (five were classified as responders at this point); 5(13% of 38) remained on the diet at the 24-month point (two were classified as responders at this point).

Table 3.3: Non-seizure-related benefits from the Ketogenic diet

Benefit	Number of cases who reported benefit, n (% of 253[cases with diet response data])
More aware/alert/brighter/attentive	n=141 (56%)
More communicative/interactive/responsive	n=61 (24%)
More vocal/verbal/improved speech	n=34 (13%)
More active	n=31 (12%)
Improved school performance	n=21 (8%)
Improved concentration	n=18 (7%)
Improved sleep	n=16 (6%)
Improved behaviour	n=10 (4%)
Steadier/better balance	n=8 (3%)
Improved eye contact	n=8 (3%)
Happier	n=7(3%)
Improved head control	n=5 (2%)
Improved motor skills	n=4 (2%)
Calmer	n=3 (1%)
Improved eating	n=2 (1%)
Improved reflux	n=1 (0.4%)

It is likely that such reported benefits are the reason why a proportion of participants remain on the diet past the 3-month point, despite being defined as non-responders in terms of seizure frequency.

3.3.5 Ketogenic diet response over time

As an example of how response to the KD can vary over time, KD responder/non-responder status over time is shown for five individuals in Figure 3.9. For some cases, responder status may be constant over time (Person 4 remained a non-responder from 3- to 18-month follow-up), whilst for others this is not the case (for example, Person 2 was a non-responder at 3-, 6- and 12-month follow-up, became a responder at 18-month follow-up and then a non-responder again at 24-month follow-up).

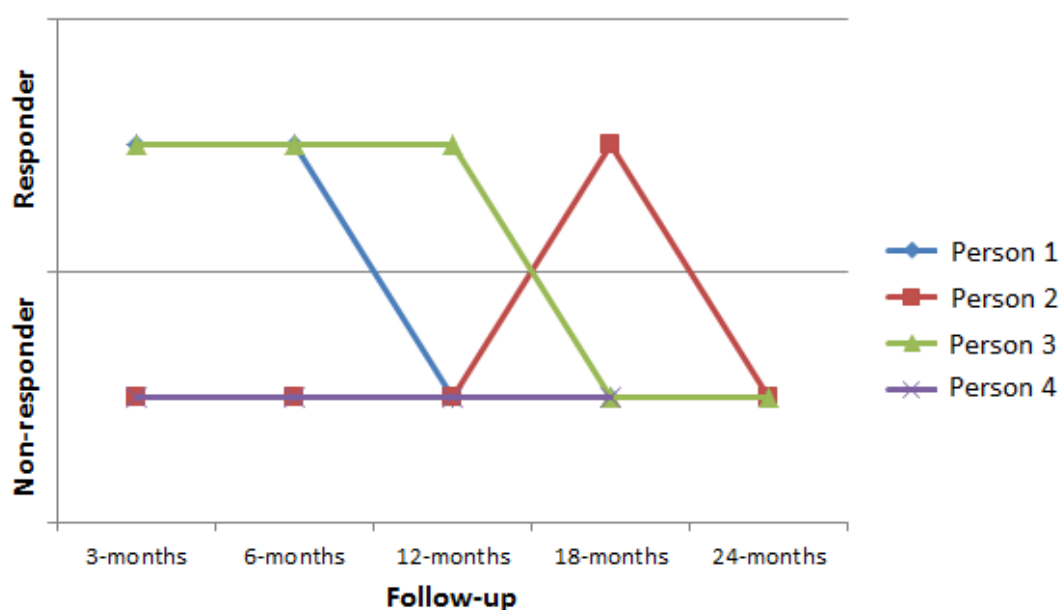


Figure 3.9: Line graph showing trend of Ketogenic diet responder/non-responder status over time for five cases

The inconsistency of KD response over time can be shown statistically. Chi-square goodness of fit tests demonstrate that response at different time points are not always correlated ($p > 0.05$), as shown in Table 3.4. Response is more likely to be correlated at 3-month and 6-

month follow-up than at later time points: the p-value for independence of response at 3-month and 6-month follow-up is far lower than for any other pairwise test.

Table 3.4: Results from Chi-square tests assessing independence of diet response at different time points

	3-month follow-up: X ² value P value	6-month follow-up: X ² value P value	12-month follow-up: X ² value P value	18-month follow-up: X ² value P value	24-month follow-up: X ² value P value
3-month follow-up	-	48.11 4.04x10 ⁻¹² *	10.52 0.009692*	4.16 0.087399	7.11 0.053992
6-month follow-up	-	-	48.86 2.04x10 ⁻⁰⁵ *	3.98 0.109783	3.19 0.160186
12-month follow-up	-	-	-	12.55 0.019647*	9.92 0.037707*
18-month follow-up	-	-	-	-	29.59 0.005534*
24-month follow-up	-	-	-	-	-

* p<0.05

When looking at KD response rates and putative factors affecting response, it may be appropriate to consider an individual's diet response over time, rather than concentrating on specific time points.

Longitudinal response was accounted for by creating a summary diet response phenotype. Some cases were easily classified as 'overall' responders or non-responders. For example, those who achieved ≥50% seizure reduction at every time point recorded were classified as responders and those who achieved <50% seizure reduction at every time point recorded were classified as non-responders. Hospital notes were used, as well as communication with patients/parents/guardians at clinic appointments and with patients' neurologists and dietitians. Individuals who consistently reported approximately 50% seizure reduction over time, or those with very fluctuating diet response over time were excluded from the summary diet response classification, as responder/non-responder status was ambiguous.

An extreme summary diet response category was also created. Extreme responders were cases who maintained $\geq 75\%$ seizure reduction at all follow-up points, or who had $\geq 75\%$ seizure reduction at all but one follow-up point (cases were not excluded if, at one time point, seizures were 50-75% reduced compared to baseline, as long as seizure frequency had reduced to $\geq 75\%$ seizure reduction by the next time point). This was to allow for a degree of variability in seizure frequency. If $\geq 75\%$ seizure reduction had been achieved but the diet was subsequently discontinued due to an increase in seizures, cases were excluded. Several cases, for example, reported $\geq 75\%$ seizure reduction at the 3- or 6-month point but then, by the next time point, seizures were either no different or were occurring more frequently compared to baseline. Cases classified as responders at the 3-month point but with no further diet response data available were excluded, as it is unknown whether they will continue to be responders at the 6-month point and beyond (this could be said for any follow-up point, as seizure frequency may vary over time, but one has more confidence that a reduction in seizure frequency is due to treatment and not due to chance if seizure reduction is consistent over more than one time points). Individuals who reported no change in seizures or increases seizure frequency at the 3-month point and subsequently weaned off the diet were classified as extreme non-responders. Individuals who reported some reduction in seizure frequency but not quite 50% reduction were excluded from this category.

The proportion of overall responders/non-responders, based on the summary response phenotype, is shown in Figure 3.10, and the extreme summary response phenotype is shown in Figure 3.11.

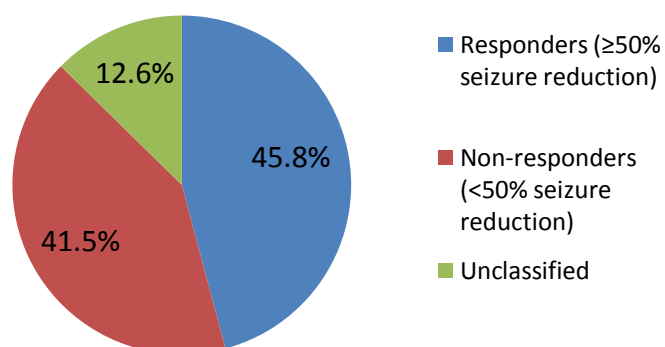


Figure 3.10: Summary diet response, n=253 (n=221 with classifiable diet response data)

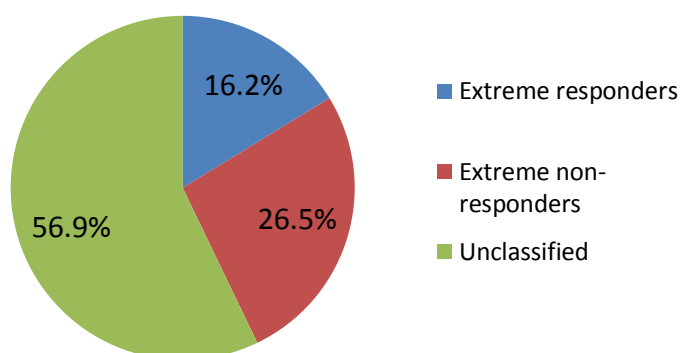


Figure 3.11: Extreme summary diet response, n=253 (n=108 with classifiable diet response data)

Using an intent-to-treat approach, 116/282(41%) would be classified as overall responders (summary diet response); 41/282(15%) would be classified as overall extreme responders (extreme summary diet response).

3.3.6 Factors affecting Ketogenic diet response

3.3.6.1 Demographic and clinical parameters

The summary response phenotypes, and 3-month KD response, were used to determine whether there was an influence of demographic/clinical factors on KD response. The 3-

month KD response cohort has the largest sample size, compared to response classified at other time points and summary response classifications, and it represents the first time point at which response to dietary treatment is formally assessed. The summary phenotypes were also used because response is not always consistent over time – the same case may be classified differently at different time points (for example, as a responder at the 3-month point, but as a non-responder at the 6-month point).

Results from logistic regression (see Table 3.5, Table 3.6 and Table 3.7) showed that the number of failed AEDs correlated (unadjusted p -value <0.05) with 3-month and summary diet response (the lower the number of failed AEDs, the increased likelihood of a favourable response) and use of topiramate correlated with both summary diet response classifications (use of topiramate was associated with increased likelihood of a favourable response). When applying a Bonferroni-corrected significance threshold, based on an alpha of 0.05 and 10 tests, no factors reached statistical significance (p -value <0.005).

Table 3.5: Putative factors affecting 3-month Ketogenic diet response

Factor	Number of responders with data available	Number of non-responders with data available	Estimate	Z value	P value
Gender	131	121	0.20	0.78	0.43
Age of seizure onset (years)	130	121	0.01	0.42	0.68
Age of diet onset (years)	131	121	-0.02	-1.22	0.23
Aetiology	131	121			
	Genetic vs Structural/Metabolic		0.05	-0.37	0.72
	Genetic vs Unknown		0.33	0.11	0.91
	Structural/Metabolic vs Unknown (using releval)		0.28	0.98	0.33
Seizure localisation	131	120			
	Focal vs Generalised		0.14	0.45	0.65
	Focal vs Focal+Generalised		-0.24	-0.75	0.46
	Focal vs Focal with secondary generalisation		0.85	0.99	0.32
	Focal vs spasms		0.34	0.36	0.72
	Generalised vs Focal+Generalised (using		-0.38	-1.14	0.25

	relevel)				
	Generalised vs Focal with secondary generalisation (using relevel)		0.71	0.82	0.41
	Generalised vs spasms (using relevel)		0.20	0.21	0.83
	Focal+Generalised vs Focal with secondary generalisation (using relevel)		1.09	1.25	0.21
	Focal+Generalised vs spasms (using relevel)		0.58	0.61	0.54
	Focal with secondary generalisation vs spasms (using relevel)		-0.51	-0.41	0.68
Number of AEDs taken at diet onset	131	121	0.09	0.69	0.49
AEDs taken at diet onset	131	121			
	Sodium valproate		0.35	1.37	0.17
	Levetiracetam		0.21	0.78	0.44
	Topiramate		0.46	1.60	0.11
	Vigabatrin		-0.18	-0.42	0.68
	Carbamazepine		-0.62	-1.31	0.19
	Lamotrigine		-0.29	-0.94	0.35
	Clobazam		0.25	0.94	0.34
Number of failed AEDs prior to diet onset	130	120	-0.10	-2.06	0.04
Diet type	130	120			
	Classical Ketogenic diet vs Modified Atkins Diet		-0.59	-1.65	0.10
	Classical Ketogenic diet vs Medium Chain Triglyceride Ketogenic diet		-0.27	-0.83	0.41
	Modified Atkins Diet vs Medium Chain Triglyceride Ketogenic diet (using relevel)		0.32	0.73	0.46
Feed	131	121			
	Oral vs Tube		0.27	0.93	0.35
	Oral vs Oral and Tube		0.53	0.99	0.32
	Tube vs Oral and Tube (using relevel)		0.26	0.45	0.65

Table 3.6: Putative factors affecting summary Ketogenic diet response

Factor	Number of responders with data available	Number of non-responders with data available	Estimate	Z value	P value
Gender	116	105	0.21	0.78	0.44

Age of seizure onset (years)	116	105	0.03	0.81	0.42
Age of diet onset (years)	116	105	-0.01	-0.63	0.53
Aetiology	116	105			
	Genetic vs Structural/Metabolic		0.01	0.20	0.84
	Genetic vs Unknown		0.01	0.27	0.79
	Structural/Metabolic vs Unknown (using relevel)		0.02	0.08	0.94
Seizure localisation	116	105			
	Focal vs Generalised		0.13	0.39	0.70
	Focal vs Focal+Generalised		-0.23	-0.66	0.51
	Focal vs Focal with secondary generalisation		0.57	0.63	0.53
	Focal vs spasms		-0.53	-0.57	0.57
	Generalised vs Focal+Generalised (using relevel)		-0.36	-1.01	0.31
	Generalised vs Focal with secondary generalisation (using relevel)		0.44	0.49	0.63
	Generalised vs spasms (using relevel)		-0.66	-0.70	0.48
	Focal+Generalised vs Focal with secondary generalisation (using relevel)		0.79	0.88	0.38
	Focal+Generalised vs spasms (using relevel)		-0.30	-0.32	0.75
	Focal with secondary generalisation vs spasms (using relevel)		-1.10	-0.87	0.38
Number of AEDs taken at diet onset	116	105	0.08	0.58	0.56
AEDs taken at diet onset	116	105			
	Sodium valproate		0.31	1.15	0.25
	Levetiracetam		0.03	1.05	0.30
	Topiramate		0.63	1.99	0.05
	Vigabatrin		-0.31	-0.69	0.49
	Carbamazepine		-0.79	-1.61	0.11
	Lamotrigine		-0.23	-0.72	0.47
	Clobazam		0.02	0.08	0.94
Number of failed AEDs prior to diet onset	115	104	-0.10	-1.93	0.05
Diet type	116	104			
	Classical Ketogenic diet vs Modified Atkins Diet		-0.37	-0.95	0.34
	Classical Ketogenic diet vs Medium Chain Triglyceride Ketogenic diet		-0.13	-0.34	0.34
	Modified Atkins Diet vs Medium Chain Triglyceride Ketogenic diet (using relevel)		0.24	0.49	0.62

Feed	116	105			
	Oral vs Tube		0.29	0.32	0.37
	Oral vs Oral and Tube		0.24	0.45	0.65
	Tube vs Oral and Tube (using relevel)		-0.05	-0.08	0.94

Table 3.7: Putative factors affecting extreme summary Ketogenic diet response

Factor	Number of responders with data available	Number of non-responders with data available	Estimate	Z value	P value
Gender	41	67	0.05	0.12	0.91
Age of seizure onset (years)	41	67	-0.02	-0.29	0.77
Age of diet onset (years)	41	67	-0.04	-1.09	0.28
Aetiology	41	67			
	Genetic vs Structural/Metabolic		-0.07	-0.10	0.92
	Genetic vs Unknown		-0.22	-0.35	0.73
	Structural/Metabolic vs Unknown (using relevel)		-0.15	-0.34	0.73
Seizure localisation	41	67			
	Focal vs Generalised		0.65	1.37	0.17
	Focal vs Focal+Generalised		0.49	0.97	0.33
	Focal vs Focal with secondary generalisation		16.44	0.01	0.99
	Focal vs spasms		n/a	n/a	n/a
	Generalised vs Focal+Generalised (using relevel)		-0.15	-0.29	0.77
	Generalised vs Focal with secondary generalisation (using relevel)		15.79	0.01	0.99
	Generalised vs spasms (using relevel)		n/a	n/a	n/a
	Focal+Generalised vs Focal with secondary generalisation (using relevel)		15.94	0.01	0.99
	Focal+Generalised vs spasms (using relevel)		n/a	n/a	n/a
	Focal with secondary generalisation vs spasms (using relevel)		n/a	n/a	n/a
Number of AEDs taken at diet onset	41	67	0.05	0.21	0.84
AEDs taken at diet onset	41	67			
	Sodium valproate		0.36	0.90	0.37
	Levetiracetam		0.06	0.14	0.89
	Topiramate		1.08	2.28	0.02
	Vigabatrin		-0.97	-1.19	0.23

	Carbamazepine		-0.80	-1.16	0.25
	Lamotrigine		-0.76	-1.47	0.14
	Clobazam		-0.07	-0.19	0.85
Number of failed AEDs prior to diet onset	41	66	-0.09	-1.14	0.25
Diet type	41	66			
	Classical Ketogenic diet vs Modified Atkins Diet		0.44	0.68	0.50
	Classical Ketogenic diet vs Medium Chain Triglyceride Ketogenic diet		-0.32	-0.58	0.56
	Modified Atkins Diet vs Medium Chain Triglyceride Ketogenic diet (using relevel)		0.12	0.15	0.88
Feed	41	67			
	Oral vs Tube		-0.62	-1.39	0.16
	Oral vs Oral and Tube		0.41	0.51	0.61
	Tube vs Oral and Tube (using relevel)		-0.22	-0.26	0.80

3.3.6.2 Epilepsy syndrome or cause

Response rates were separated out according to groups of patients with the same epilepsy syndrome or aetiology. Those with epilepsy with myoclonic atonic seizures had higher response rates than expected (considering response rates in the cohort as a whole, although no statistical analyses were performed, as the sample size in each group was small; most participants do not have an epilepsy syndrome diagnosis or have epilepsy of unknown aetiology): 12/14(85.7%) cases with epilepsy with myoclonic atonic seizures groups were classified as responders at the 3-month point, and 10/14(71%) were classified as responders according to the summary diet response phenotype (one case was excluded from the summary response classification due to variable seizure frequency over time), five (36% of 14) of which were extreme responders.

Response rates in other groups (those including ≥ 10 cases) are as follows:

- i) 7/15(47%) cases with Dravet syndrome (14 of which had a known mutation in *SCN1A*) were responders at the 3-month point; 7/15(47%)

were overall responders (summary diet response), of which three (20 % of 15) were extreme responders.

- ii) 10/15(62.5%) cases with West syndrome were responders at the 3-month point (another patient with West syndrome was recruited [n=16] but KD response was unclassifiable in this case), and 8/15(53%) were overall responders (summary diet response), three (20% of 15) of which were classified as extreme responders.
- iii) 6/13(46%) cases with Lennox-Gastaut syndrome were responders at the 3-month point; 3/13(23%) were overall responders (summary diet response), of which one (8% of 13) was an extreme responder.
- iv) 11/20(55%) cases with epilepsy due to hypoxic ischemic encephalopathy were responders at the 3-month point; 11/20(55%) were overall responders (summary diet response), of which seven (35% of 20) were extreme responders.
- v) 5/10(50%) cases with epilepsy due to polymicrogyria were responders at the 3-month point; 4/10(40%) were overall responders (summary diet response), of which one (10% of 10) was an extreme responder.

3.3.6.3 Biochemical parameters

Possible associations of selected biochemical parameters measured at 3-month, 6-month and 12-month clinic appointments with diet response at the corresponding follow-up time points were investigated by logistic regression, as shown in Table 3.8, Table 3.9 and Table 3.10 respectively. The association of biochemical parameters was not assessed at the 18- or 24-month points, due to small sample sizes.

A Bonferroni-corrected significance threshold, based on an alpha of 0.05 and 15 tests, was imposed. No factors reached statistical significance when adjusting for multiple testing (p-value<0.003).

Table 3.8: Putative factors affecting diet response at 3-month follow-up: biochemical parameters

Biochemical parameter, measured in blood	Estimate	T value	P value	Number of cases with data (3-month diet response and biochemical measure)
Free carnitine (μmol/L)	0.007	2.43	0.02	104
Acetyl carnitine (μmol/L)	0.01	2.19	0.03	93
Propionyl carnitine (μmol/L)	-0.01	-0.09	0.93	76
Butyryl carnitine (μmol/L)	-0.44	-1.15	0.25	76
Isovaleryl carnitine (μmol/L)	0.09	0.17	0.86	76
Hexanoyl carnitine (μmol/L)	0.17	0.18	0.86	68
Octanoyl carnitine (μmol/L)	0.36	0.64	0.52	76
Tetradecenyl carnitine (μmol/L)	0.62	0.88	0.39	61
Palmitoyl carnitine (μmol/L)	0.25	2.01	0.04	67
Creatinine (μmol/L)	0.0009	0.29	0.77	146
Glucose (mmol/L)	0.02	0.20	0.84	86
Non-esterified fatty acids (mmol/L)	-0.01	-0.11	0.92	111
Beta-hydroxybutyrate (mmol/L)	0.04	1.29	0.20	122
Cholesterol (total) (mmol/L)	-0.003	-0.09	0.93	129
Triglycerides (mmol/L)	-0.01	-0.42	0.68	128

Table 3.9: Putative factors affecting diet response at 6-month follow-up: biochemical parameters

Biochemical parameter, measured in blood	Estimate	T value	P value	Number of cases with data (6-month diet response and biochemical measure)
Free carnitine (μmol/L)	0.0008	0.25	0.80	77
Acetyl carnitine (μmol/L)	0.007	1.54	0.13	59
Propionyl carnitine (μmol/L)	-0.13	-1.16	0.25	53
Butyryl carnitine (μmol/L)	-0.20	-0.37	0.71	53
Isovaleryl carnitine (μmol/L)	1.38	1.81	0.08	52
Hexanoyl carnitine (μmol/L)	-0.53	-0.53	0.60	53
Octanoyl carnitine (μmol/L)	-0.99	-1.30	0.20	49
Tetradecenyl carnitine (μmol/L)	1.37	1.64	0.11	49

Palmitoyl carnitine (μmol/L)	-0.03	-0.18	0.86	52
Creatinine (μmol/L)	0.0002	0.04	0.97	91
Glucose (mmol/L)	-0.07	-0.83	0.41	71
Non-esterified fatty acids (mmol/L)	-0.15	-0.76	0.45	67
Beta-hydroxybutyrate (mmol/L)	0.03	1.03	0.31	70
Cholesterol (total) (mmol/L)	0.06	1.45	0.15	87
Triglycerides (mmol/L)	0.006	0.19	0.85	87

Table 3.10: Putative factors affecting diet response at 12-month follow-up: biochemical parameters

Biochemical parameter, measured in blood	Estimate	T value	P value	Number of cases with data (12-month diet response and biochemical measure)
Free carnitine (μmol/L)	-0.001	-0.32	0.75	44
Acetyl carnitine (μmol/L)	-0.005	-0.94	0.35	39
Propionyl carnitine (μmol/L)	0.10	0.43	0.67	33
Butyryl carnitine (μmol/L)	0.36	0.79	0.44	33
Isovaleryl carnitine (μmol/L)	0.53	0.62	0.54	33
Hexanoyl carnitine (μmol/L)	-1.24	-0.82	0.42	33
Octanoyl carnitine (μmol/L)	-2.39	-2.81	0.01	33
Tetradecenyl carnitine (μmol/L)	-0.13	-0.17	0.87	33
Palmitoyl carnitine (μmol/L)	-0.08	-0.45	0.65	33
Creatinine (μmol/L)	-0.003	-0.75	0.46	61
Glucose (mmol/L)	-0.18	-1.76	0.09	41
Non-esterified fatty acids (mmol/L)	0.11	0.79	0.44	52
Beta-hydroxybutyrate (mmol/L)	-0.02	-0.58	0.56	56
Cholesterol (total) (mmol/L)	-0.02	-0.44	0.66	54
Triglycerides (mmol/L)	0.04	0.82	0.42	53

3.3.7 Use of phenotypic data for genetic analyses

Seizure frequency will be the only parameter used to classify individuals as responders/non-responders for use in genetic analyses. This is not because other effects of dietary treatment are not considered clinically meaningful, but because they have not been

quantified in this study and therefore, classification of KD response that includes these parameters could not be reliably replicated in other cohorts.

If there is there a genetic basis to KD response, it is unknown what architecture it takes.

Different mechanisms may influence KD response at different time points, or (as in the case of *SLC2A1* mutations), certain variants may influence KD response, no matter whether this is in the short- or long-term. Due to this uncertainty (this is an exploratory study), the following phenotypes will be used for genetic analyses:

i) 3-month KD response

This represents the largest cohort and so the maximum proportion of available data is used. Furthermore, in clinical practice, response to the KD is commonly judged at 3-month follow-up. Based on this judgement, a decision is taken whether to discontinue or continue treatment (if continued, the benefits of the KD are weighed up against any disadvantages, at each time point). Using 3-month KD response as a phenotypic measure is thus clinically useful.

The number of individuals with diet response data falls over time (many people have weaned off the diet). Due to the small sample size, at least in terms of genetics studies, the 3-month response cohort only provides limited power to detect an effect of the factor being investigated; the 6-, 12-, 18- and 24-month cohorts have even less power and non-responders are under-represented at later time points. For these reasons, separate analyses will not be conducted using response at these time points, although they are used to classify individuals as overall responders and non-responders.

ii) Summary KD response and extreme summary KD response

Using summary response classifications accounts for the fact that response at specific time points may not be representative of an individual's overall experience with the KD (an individual may be classified as a responder at 3-month follow-up but as a non-responder at 6-months). In this way, associations with genes affecting KD response in the broader sense are searched for.

The summary diet response phenotype will be used for analyses investigating the putative influence of common genetic variation, which are assumed to have a small phenotypic effect. The extreme summary diet response phenotype will be used for analyses investigating the putative influence of rare genetic variation, which are thought, in general, to have a larger phenotypic effect compared to common variation.

Comparisons may be made with 3-month KD response analyses; some genes may influence KD response in the short-term, whereas others influence response in the long-term, potentially due to longer-term gene expression changes.

One individual was found to have GLUT1 deficiency after completion of this chapter (identified from whole exome sequencing) and so was excluded from all genetic analyses.

3.4 Discussion

Key findings:

- All phenotypic data (except for compliance questionnaires) used in this study were taken from sources designed for clinical monitoring, not for research purposes.
- Recognising the limits of the data available, the demographics of this cohort (predominantly young people with severely drug-resistant epilepsy), and KD response rates are similar to those reported in the literature.

- Compliance to the KD was assessed in a proportion of patients with use of questionnaires and serum BHB levels taken at clinic appointments. Most patients were thought to be compliant, but more objective measures of assessing compliance, which are more representative of the whole period of time between clinic appointments, are needed.
- Seizure frequency was used to measure KD response. This was the only effect of the diet to be quantified (at least, to the extent possible when data is taken mostly from clinic letters). Response may be classified at specific follow-up points, or as a summary of response over time.
- Other parameters, such as a reduction in seizure severity or increased alertness, are often clinically relevant to individuals following the KD – people may remain on the diet despite <50% reduction in seizure frequency. These parameters are not routinely quantified as part of routine clinical monitoring and so were not used to classify KD response in this study. Tools, such as seizure severity scales or quality of life questionnaires, may be useful in further studies.
- None of the non-genetic factors evaluated were significantly associated with KD response (either at 3-month follow-up or when using summary diet response phenotypes) after adjusting for multiple testing.

3.4.1 Cohort demographics

The demographics of this cohort, which has one of the largest sample sizes reported, are typical of many KD groups presented in the literature. There is a tendency for patients to have an early age at seizure onset and diet initiation (this is also affected by the small number of centres offering KD treatment for adults), have failed multiple AEDs, and they often have some level of developmental delay. Many have unknown causes of epilepsy or a clinical presentation that does not easily fit into a particular epilepsy syndrome.

The difficulties in accurately phenotyping patients with epilepsy must be highlighted. Determination of seizure localisation, for example, even when based on EEG data and clinical seizure manifestation, may be open to interpretation. In these cases, one must trust in the experts (such as neurophysiologists) who interpreted the data in the first place. Patients are often classified under such umbrella terms as Idiopathic Generalised Epilepsy, particularly in older clinic letters, thus heterogeneity within groups may be wider than expected. When adopting a relatively new classification system, as is the case here, some interpretation is often required when extracting information from sources that have used previous systems.

Even with deep phenotyping, some desired data will unavoidably be missing. There is, for example, (conflicting) evidence that measures of BMI are correlated with KD response; recordings of weight, weight-for-height z-score and BMI at diet onset, or changes in these variables during dietary treatment, could not be included here due to incompleteness and inconsistencies in the data. In most cases, patients' weight and height were measured at diet information clinic sessions. However, the diet may not have been initiated until much later (over a year in some cases) and all measurements may not have been recorded again immediately before commencement of the diet. Missing values arose from unattended appointments or appointments when measurements were not taken. When considering growth in paediatric patients, such inconsistencies would make inter- and intra-patient comparison problematic. As many data sources were consulted as possible to avoid missing data for other variables.

3.4.2 Compliance

Compliance questionnaire scores indicated high compliance to the diet in approximately three quarters of patients for whom questionnaire data were available, although the definition of high compliance used here, which is based on the frequency distribution of

the data, might change if more questionnaires were collected. Notwithstanding the fact that compliance data are not available for all cases, scores calculated from the questionnaires assume, on the part of the patient or parent/guardian(s), complete honesty and a thorough understanding of the restrictions of the KD and of the questions being asked. Although it has been reported that people who admit non-compliance to medication tend to be telling the truth [456-458], indirect methods of measuring compliance, such as questionnaires, may lead to over-estimation of compliance levels [459]. Reactions to the statement 'We alter the diet' may depend on the type of KD followed, how often changes are made by the patient's dietitians, and whether such changes were accounted for here. Despite the advantages of basing the questionnaire on a validated one, perhaps further adjustments could have made it more diet-specific.

The moderate Cronbach's alpha correlation coefficient may have been affected by the small number of items in the questionnaire, the positive distribution of scores or the small sample size. As a non-normal distribution of scores reduces the precision of the coefficient, underestimating the population reliability, a larger sample size (>100 for a moderate correlation coefficient or >1000 for small correlation coefficient) would help approximate the non-normal distribution towards normality and reduce the negative bias imposed on the alpha value [453]. This has been found to reduce bias to a greater extent than when the number of items/test length was increased.

An alternative method of assessing reliability of the compliance questionnaire, 'test-retest', would have provided an indication of test reliability over time. Employing both Cronbach's alpha and the 'test-retest' method to assess reliability of the compliance questionnaire may have yielded more easily interpretable results. The likelihood that participants would have completed and returned the same questionnaire on more than one occasion, however, is low.

Used as a measure of diet compliance, serum BHB levels are also hard to interpret. Although generally considered more accurate than urine ketones, which are influenced by fluid intake, urine volume and concentration [460], serum BHB may be affected by other medications, recently-completed exercise, constituents of recent meals/snacks, and time of day. Inter-individual variability in ketone body production and maintenance is high [165] and measures taken on one day may not be representative of the whole period of time between each follow-up point. It would have been preferable to use measures taken at home by parents on a twice-daily basis. However, whilst most parents claim to take these measurements, at least when dietary treatment has been recently commenced and dietary manipulation is still occurring, urine ketones are measured in most children (serum BHB would be preferable for research purposes) and these data were seldom available for the 28 days prior to each follow-up point, when seizure frequency was determined. It is becoming increasingly common for parents to hand over 'diaries' of ketone levels measured at home over a prolonged period of time, although this varies among centres. If all participants were to be recruited prospectively, a higher proportion of these data may be obtained. Measurements taken by parents may be subject to a greater level of error than measurements taken in hospitals, but use of repeated measurements should be more representative of the period of time between each follow-up point. Measurement of breath acetone levels is an alternative way to assess ketosis on a regular basis (levels correlate with urinary and serum ACA and serum BHB [376, 461]), but the equipment needed to do this is not commonly available outside a laboratory setting.

Reports from the literature assessing KD compliance have so far relied on urine and blood ketone testing [106], where 3/12(25%) patients were described as having 'none' or 'mild' compliance, or on communication with parents/patients [73], where 12/56(21.4%) patients had 'poor' compliance. It is unknown what threshold was adopted to define compliance

levels defined by ketone body measurements and so it is difficult to compare compliance levels in this cohort (also those derived from the questionnaires) to anything in the literature.

Based on the fact that compliance data are not available for all participants, together with the realisation that many other (uncontrollable) factors may have influenced results, no participants were excluded from analyses based on compliance measures. One must be aware of potential misclassification bias if some participants were incorrectly classified as non-responders simply because they were non-compliers. Other methods, such as weighed food records, would have more accurately ensured that all participants were compliant, but this would have been difficult to achieve with a large sample size and would have required all participants to be recruited prospectively. The compliance data collected do, at least, provide some assurance that most patients were actually following the diet, something that is often neglected in drug response studies. They do, at least, provide some assurance that most patients were actually following the diet, something that is often neglected in drug response studies.

3.4.3 Diet response

Responder rates in this cohort are consistent with those reported in the literature. The proportion of responders increases as time goes on, which is consistent with the fact that non-responders are less likely to remain on the diet long-term. The proportion of responders may be higher than those reported in prospective studies with an intent-to-treat analysis; in this study, even if all participants who were approached for recruitment were included when calculating responder rates, a true intent-to-treat approach could not be used as some participants were recruited retrospectively. However, one would not expect responder rates to be too inflated, as the majority of participants have been recruited from tertiary referral centres, where particularly severe cases of intractable

epilepsy are seen. Only patients with hospital follow-up appointments could be recruited retrospectively, which may have resulted in a bias towards cases with multiple co-morbidities, which may be harder to treat.

Several difficulties with classifying response to the diet must be considered. For example, one is reliant on the patient or his/her parent/guardian/carer(s) for the recording of seizure frequency; this may lead to under- [462] or over-reporting [463], depending on the motivation, capability and priorities of the recorder. The proportion of participants with completed seizure diaries at baseline and four weeks prior to the next follow-up point was low, depending on the KD service protocol in each hospital, parental/patient commitment and whether the participant had previously taken part in research or schemes for which recording of seizures was required. This proportion decreased with time. Clinic letters, which essentially provide a second-hand report of information from patients/parents/guardians, were the most common source used to ascertain response to the diet; created primarily for patient care and administrative purposes (not for research) [464], they do not guarantee accuracy or reliability. Some cases remained unclassified due to inadequate documentation. For example, comments such as seizures were 'better' or 'reduced' were deemed inadequate to determine response at that time point. In some cases, the most recent pre-diet clinic appointment was a substantial amount of time before the commencement of dietary treatment, for example due to the wait for funding approval or medication changes, and so baseline seizure frequency could not be ascertained. Classifying diet response was particularly difficult when patients were under the care of another centre when they started the KD, often quite some time prior to recruitment - obtaining hospital records from a site not directly involved in the study was problematic and so details regarding diet response were insufficient for accurate classification. Personal communication between the researcher and patients/parents/guardians and participants'

consultants and dietitians aided response classification, both for cases recruited prospectively and retrospectively. The researcher was better able to judge, for example, whether the description of seizure frequency at each time point was accurate, depending on the reporter, seizure type, or seizure frequency itself (if the patient is having hundreds of seizures a day, it may be impossible to obtain an accurate number).

More widespread use of seizure-diaries may further improve the data quality and provide documentation of the frequency of each seizure type, which the resolution of the data from clinic letters did not allow. In the future, it may be feasible that completion of seizure diaries for at least 28 days prior to starting the diet, and 28 days prior to each follow-up point, is stipulated as a requirement for involvement in the study, at least for participants recruited prospectively. Whereas parental reporting of seizures (for example, from phone interviews and mailed questionnaires) can poorly correlate with data from medical records [465], through use of diaries, parental reporting has been shown to have high concordance with objective methods (such as an actigraph to measure sleep schedules) [466]. However, considering that the act of observing or self-observation itself introduces bias, influencing the phenomenon being observed, seizure diaries may not be a panacea (although no reports were found that show the extent of bias introduced specifically by seizure diaries). Accurate reporting depends on awareness of seizures (over half of seizures, in particular focal seizures, may not be recognised by the patients themselves or carers [462, 467-469]) and the motivation and capabilities of the recorder. Recall bias may affect accuracy if the recorder forgets to input an entry around the time of an event. Electronic diaries and supervised diaries, where information inputted by the recorder is reviewed by a researcher, may lead to more consistent data entry, as recorders are reminded to fill in the diary, and potentially increase diary validity [470]. Having said this, for individuals whose seizures occur in weekly/monthly clusters, seizure reduction on the diet may not be accurately

reflected in seizure diaries (the cluster may occur just before the 28 days prior to baseline and so, according to the seizure diary, the patient would be seizure-free before starting the diet) and so clinic letters may be more appropriate for ascertaining diet response. For individuals with non-clinical seizures, even the use of seizure diaries may not help parents/guardians to ascertain seizure frequency; frequency counts may only be possible with an EEG recording. In this study, such cases were unclassifiable as responders or non-responders.

Even if the number of seizures documented were 100% precise, the pattern of seizure frequency may simply reflect the natural evolution and fluctuation of the seizure disorder. One may question whether changes in seizure frequency were due to initiation of the KD or whether they would have happened anyway and subsequently returned to baseline – regression to the mean [471]. One way to avoid such doubt and account for recent fluctuations in seizure frequency would be to take seizure-freedom (potentially measured by the ILAE definition of drug resistance [7] or ‘Rule of Three-to-Six’ [472]) as the measure of diet response. It has been reported that only seizure-freedom, or >90% seizure reduction [473] will result in measureable improvement in quality of life in adults [474], with seizure-free adults reaching similar levels to those of the general population [475]. However, while it is true that seizure-freedom (or as close to seizure-freedom as possible) would be ideal, and has been associated with improved quality of life in children compared to those with persistent seizures [476], the probability of achieving this with any new treatment in a cohort with drug-resistant epilepsy is minimal [477, 478]. A $\geq 50\%$ seizure reduction, a standard measure of treatment effectiveness used in clinical trials [479], was viewed as clinically useful in this cohort. An observed improvement in seizure control, defined as $\geq 50\%$ seizure reduction, from use of everolimus was accompanied by improved quality of life in children [480]. A $\geq 50\%$ fluctuation in seizure frequency is also less likely to occur due

to chance, compared to lesser degrees of fluctuation [481]. Calculating seizure frequency in a defined epoch prior to baseline and each follow-up point seemed more appropriate than methods that rely on previous seizure-free intervals, as, in view of the severity of the condition of most participants (seen in tertiary, sometimes considered quaternary, care), many may continue to have daily seizures even after a dramatic seizure reduction. Clinic letters may also not provide detailed or accurate data regarding the longest seizure-free interval in the previous 12 months.

By using repeated measures of response over time, confidence is further increased that changes in seizure frequency were due to dietary treatment and not the natural evolution of the seizure disorder. As shown by chi-squared tests, 3-month diet response is likely to correlate with 6-month or (to a lesser extent) 12-month diet response, but not with 18- or 24-month response. Whilst calculating seizure frequency at various time points compared to baseline is another approach to classifying KD response, which may be useful in genetic analyses to ascertain whether different genes influence response at different time points, it is clear that response taken at one time point may not accurately reflect response at a later time point. Summary response phenotypes account for this variability in response over time. Difficulty arises with individuals who have been following the diet for a short period of time. As shown by individuals who have followed the diet long-term, response, for example, at the 3-month point, may not correlate with response at later time points. Thus, summary phenotypes for cases with data at few time points may not be accurate or may differ if more data were to be collected in the future. For this reason, in the extreme summary diet response category, individuals with only 3-month response data were excluded (unless, of course, they weaned off the diet due to lack of response). A disadvantage with this category is that the sample size is dramatically reduced and so the power to detect an association between response and demographic/clinical factors is low.

As this category includes only the most extreme responders and non-responders, however, it should have the most power to detect associations with rare variants with a large effect size on KD response.

3.4.4 Factors affecting Ketogenic diet response

Out of all factors assessed, the number of failed AEDs prior to diet onset and use of topiramate at diet onset were most closely associated with KD response. They did not reach statistical association when correcting for multiple testing, although it must not be forgotten that the Bonferroni adjustment is conservative; whilst it decreases the chance of committing a Type I error, it increases the chance of Type II error [449]. Perhaps with a larger sample size, these, or other factors would reach statistical significance after adjustment for multiple testing.

The correlation detected between number of failed AEDs and KD response (the greater the number of AEDs, the less likely a favourable response to the diet) corresponds with two reports in the literature [105, 482], although other evidence is conflicting [55, 118, 121, 123, 124, 127, 208, 360]. Patients are sometimes advised to wean off topiramate before initiating dietary treatment due to the increased risk of metabolic acidosis and renal stones, both from the drug [483] and the KD [484-486]. However, according to the literature, responder rates in patients taking topiramate and following the KD are similar to responder rates in patients taking other AEDs [211, 487]. Number of failed AEDs prior to diet onset is likely to be a surrogate for disease severity (individuals with more severe epilepsy will probably have tried, and failed, more AEDs than those with a less severe condition), and could be influenced by age, so may not be worth including as a covariate in the genetic analyses. Use of topiramate may also be inextricable from other factors, as this AED is often used for patients with Lennox-Gastaut syndrome [483], who are particularly refractory to treatment but may respond more favourably to the KD than patients with other epilepsy

syndromes [113, 359]. For this reason, together with the fact that these factors did not reach statistical significance when correcting for multiple testing, these factors will not be included as covariates in genetic analyses.

Considering the conflicting evidence presented in Chapter 1, the lack of effect of other factors on KD response in this cohort is unsurprising. Some factors were not suitable for statistical analysis, such as epilepsy syndrome, as there were too few cases in each category, and intellectual status, as all but 18 cases had some degree of developmental delay. Particularly favourable response rates in patients with epilepsy with myoclonic atonic seizures are consistent with reports from the literature [338-340, 359]. *SLC2A1* mutations have been detected in approximately 5% of patients with this condition [348].

No biochemical parameter had a significant relationship with KD response at all time points investigated. One may expect serum BHB, or even glucose, to relate to diet response, although evidence from the literature is conflicting. Due to their role in fatty acid transport, it is of interest that various carnitines were associated ($p\text{-value} < 0.05$) with diet response at 3-month follow-up, although (with the exception of octanoyl carnitine at 12-month follow-up) these relationships were not maintained over time and, when adjusting for multiple testing, statistical significance was not reached. Blood is only taken (except in emergency situations) when patients attend KD clinic appointments and it must be considered that measures taken on one day may not be representative of the whole 3-month period. Urine, or sometimes blood ketones may be measured on a daily basis, particularly at the start of treatment. If these data were readily available from hospital records, it could be better determined whether levels correlate with seizure fluctuation over time.

3.4.5 Factors beyond seizure frequency

Although seizure frequency is the most common measure of response to the KD in the literature, other (potentially clinically useful) aspects of seizure control, such as seizure

severity and duration, or quality of life, must not be disregarded. Non-responders to the diet in terms of seizure reduction are not exempt from gaining other benefits on seizures, cognition or behaviour. The fact that a proportion of cases who fail to achieve $\geq 50\%$ seizure reduction stay on the diet past the initial 3-month trial period demonstrates that other parameters are clinically relevant for some patients and their families and may outweigh any disadvantages that adherence to a dietary regime entails. A reduction in seizure severity and quicker recovery time have been shown to lead to improved quality of life in people with epilepsy [488-491]; thus, it is not surprising that individuals may remain on the KD if they obtain such benefits, even if they do not achieve $\geq 50\%$ seizure reduction.

Improved cognition may influence how long somebody remains on the KD more than seizure control [147]. Increased alertness is another commonly reported benefit from the diet [78, 143, 492, 493], as in this cohort, although the proportion of people who reap such benefits is not consistently reported in the literature and no reports were found demonstrating that this factor influences duration of the KD more than seizure frequency.

The number of patients said to have successfully reduced AEDs in other studies are generally consistent with this cohort [358, 494], although proportions vary widely, depending on the population. For example, in a cohort of patients with Dravet syndrome, all those who did not achieve a dramatic reduction in seizures with the KD managed to reduce the number of AEDs taken to one or two [347]. Of course, it may be difficult to gauge how an individual reacts over a long period of time to a reduction in AEDs, due to the fluctuating nature of epilepsy.

The approach used here – recording whether patients or his/her parent/guardian(s) reported any other effects of the diet aside from seizure frequency – is somewhat rudimentary and subjective. Although as many data sources as possible were used, this information was not systematically reported or recorded. In order to incorporate such

measures when classifying KD response and to allow replication of results by external groups, these parameters need to be quantified. Regarding effects on seizures other than frequency alone, quantifiable scales, such as the Chalfont [495], National Hospital [496] and Liverpool [497] seizure severity scales, would be possible tools. Use of a seizure frequency and severity scale, such as one based on the Veterans administration scale [498], would allow for consideration of more than one component of seizure control. Other benefits, such as alertness or happiness, may be more problematic, as different measures would be best assessed with different tools. Perhaps a quality of life questionnaire, such as the Cambridge London QoL questionnaire, currently being developed for use in KD clinics, would best encompass a range of such measures. With this range of tools, a more accurate and comprehensive classification of KD response can be made. This would be particularly important for cases who achieve approximately 50% seizure reduction (borderline cases) or when diet response fluctuates over time.

3.4.6 Conclusion

This cohort appears representative of the KD population as a whole, with response rates and cohort demographics consistent with the literature. Proportions of responders and non-responders, however, are highly dependent on definitions of diet effectiveness.

Unlike many studies, where compliance to the diet is simply assumed or based solely on ketone levels, in this cohort (where data were available), serum BHB and questionnaires were used to demonstrate high compliance. However, serum BHB may have been influenced by other factors and so caution must be taken. An adjusted questionnaire with improved internal consistency is worth consideration for the future.

When investigating putative correlations between diet response and patient demographics or clinical/biochemical data, response at all follow-up points available were considered. In this cohort, number of failed AEDs and use of topiramate showed strongest evidence

(although not statistically significant when adjusting for multiple testing) for association with diet response, but the extent to which this is influenced by other factors, such as disease severity, epilepsy syndrome and age, is unknown. The association between various carnitine fractions and diet response at various time points may be worth investigating in a larger cohort.

Response may be classified at specific time points or as a summary of response over time. There is no single 'correct' phenotype; the criteria used to best classify response depends on what the data is being used for (for example, which genetic model is considered). Sole use of quantitative measures of response – namely seizure frequency – reduces ambiguity and facilitates replication of results, but it fails to account for other, potentially clinically-relevant measures of response, such as seizure severity or increased alertness. In future studies, quantitative measurement (where possible) of such parameters would ideally be included. A uniform definition of KD response would be needed for replication of genetic analyses.

4 Candidate gene analysis

4.1 Introduction

4.1.1 The candidate gene approach

When exploring the genetics behind a complex trait such as KD response, the candidate gene approach is a powerful and economical method of analysis [499]. Candidate gene studies are based on *a priori* knowledge about genes/variants and their effects on biological pathways or the trait in question. Whilst this reliance may, at first, appear limiting (as other biological pathways not previously associated with a phenotype may be of equal or more importance than those already identified), candidate gene studies have contributed much to our understanding of nutrigenetics [500]. One example is the effect of the homozygous *APOE4/E4* genotype on response (in terms of reducing plasma cholesterol and cardiovascular disease risk) to changes in dietary intake of total fat, cholesterol and fatty acid composition [501]. Some results have even had clinical utility; for example, testing patients for *KIF6* status prior to starting statin treatment was associated with improved treatment adherence at the 6-month point [502].

For genetic association studies (whether the analysis includes variants from specific genes or from across the whole genome), single nucleotide polymorphisms (SNPs - the substitution of a single nucleotide for another in a specific genetic location) are usually the marker of choice. They occur throughout the genome, more frequently in non-coding than in coding regions [503], and are considered to be a major genetic source of phenotypic variability within a given species [504]. The inclusion of all variants found from sequencing, rather than prior selection of polymorphisms, allows consideration of variants with lower MAF, although large sample sizes are needed to conduct association studies with rare variants. SNPs may be grouped into haplotype blocks, where variants are tightly linked with

each other, and these blocks used as markers for association studies. Haplotype markers are usually inferred, although blocks may differ, not only between loci but also between populations [505]; it is possible to directly infer haplotypes using molecular haplotyping methods [506], but these are rarely used in practice as they are expensive and time-consuming. It has been said that haplotypes provide more power than single SNPs in the presence of multiple susceptibility alleles at a locus [507], and that haplotypes can capture epistatic interactions between SNPs [508], but evidence is conflicting and it seems that the putative advantages of haplotype analyses are dataset-dependent [509].

Phenotypes may be quantitative, such as biochemical parameters, or binary, where one group of individuals ('cases') may be affected by the disease in question, or have responded in a particular way to some treatment, and the other ('controls') may be unaffected, randomly selected from the population, or have responded differently to treatment. An increased frequency of a particular SNP allele in cases compared to controls suggests an association of the allele with increased risk of disease or (un)favourable response to treatment. With quantitative phenotypes, a unit change in the measure in question may be associated with a SNP allele or a particular genotype.

A major source of confounding in genetic association studies can be due to differences in allele frequencies between case and control groups arising from different populations, rather than the phenotypic trait in question. If population groups are not equally distributed between cases and controls, population-specific SNPs may appear to be associated with the phenotype simply because of differences in allele frequency in the various groups amongst cases and controls [510], or a lack of effect in one population compared to another [511]. In candidate gene studies, if genome-wide data are not available to calculate genetic distance/relatedness between participants, recruited

individuals should ideally be as ethnically-homogenous as possible, or self-reported ethnicity should be used as a covariate in analyses.

Candidate gene analyses have a reduced multiple testing burden compared to its association study counterpart, the genome-wide association study (GWAS), and thus often boast higher statistical power to detect associations [512]. This is especially important with small sample sizes, although variant effect size and allele frequency will also affect the power a study has to detect an association [513].

In this study, candidate gene variants were identified by Sanger sequencing. Although more time-consuming and costly compared to next-generation sequencing methods (which will be explored in more detail in subsequent chapters), Sanger sequencing is considered the 'gold standard' [514] and remains in use predominantly for candidate gene analyses, small-scale studies, or to verify variants identified from next-generation sequencing methods. The accuracy of base-calling is generally higher with Sanger sequencing than with next-generation sequencing methods [515] and incomplete representation and coverage with next-generation sequencing may lead to mutations being missed [516, 517]. Sanger sequencing is based on a chain-termination method, whereby the DNA is denatured to give single strands, a specially-designed primer is added and binds to the DNA sequence of interest, and different dideoxynucleotides (which are the same as the nucleotides in DNA, except for the fact they contain a hydrogen group on the 3' carbon instead of a hydroxyl group) are integrated into the sequence. When the enzyme DNA polymerase is added, the new strand of DNA is synthesised and when a dideoxynucleotide is incorporated into the chain instead of a normal nucleotide, the chain is terminated. This produces DNA sequence reads of varying lengths. The final products are run on a gel and separated by electrophoresis, which orders the DNA fragments by size; depending on whether the primer or nucleotides were radioactively or fluorescently labelled, the gel is exposed to

either ultraviolet light or x-ray in order to visualise the DNA bands. Alternatively, each dideoxynucleotide can be labelled with different colour dyes, which each fluoresce at a different wavelength, so when they are run on a gel, the identity of each band can be determined according to the wavelength at which it fluoresces.

4.1.2 Candidate genes: background information

Two genes have recently been suggested as attractive candidates putatively implicated in the antiepileptic mechanisms of the KD, *Kcnj11* and *Bad* [226]:

- i) *Kcnj11* (inwardly-rectifying potassium channel, subfamily J, member 11) encodes the Kir6.2 pore-forming subunit of K_{ATP} channels.

K_{ATP} channels are best known for their role in glucose-stimulated insulin-secretion from pancreatic β cells: when glucose enters the β cell and is metabolised, the increase in intracellular ATP and decrease in magnesium-ADP inhibit the K_{ATP} channel and cause the channel to close. This leads to membrane depolarisation, which activates voltage-gated L-type Ca²⁺ channels, causing insulin to be secreted from the cell by exocytosis [518, 519].

Akin to their role in maintaining glucose homeostasis in pancreatic β cells, evidence suggests that Kir6.2-containing K_{ATP} channels are essential for glucose responsiveness in glucose-responsive neurons located in the ventromedial hypothalamus: approximately a quarter of *Kir6.2*^{+/+} neurons displayed increased spontaneous discharge rate in response to increased extracellular glucose concentrations; *Kir6.2*^{-/-} neurons already showed a higher discharge rate at lower glucose levels compared to controls, but no further increase was observed when the glucose concentration was increased [520].

The Kir6.2 subunit of K_{ATP} channels has a suppressive effect on substantia nigra pars reticulata neuronal activity and determines seizure threshold during brief hypoxia: although the firing rate of *Kir6.2*^{-/-} neurons did not differ significantly from wild-type neurons in the

normal state, in brief (90 seconds) hypoxia, the firing rate decreased in wild-type but increased in *Kir6.2*^{-/-} neurons; wild-type mice showed no response to brief (150 seconds) hypoxia caused by oxygen deprivation, while all *Kir6.2*^{-/-} mice responded with a myoclonic jerk, followed by a severe tonic-chronic convulsion and death [521].

The E23K (p.Leu270Val) polymorphism in *KCNJ11* has been associated with Type II diabetes, although evidence is inconsistent [522, 523]. Inactivating mutations in the gene lead to familial persistent hyperinsulinaemic hypoglycaemia of infancy (PHHI) [524] and heterozygous activating mutations are a common cause of permanent neonatal diabetes mellitus (PNDM), with a prevalence of 33-50% in large series of patients [519, 525]; some patients with such mutations also suffer from developmental delay, muscle weakness and epilepsy. *KCNJ11* mutations have been identified in individuals with severe, intermediate and mild forms of DEND syndrome (developmental delay, epilepsy, and neonatal diabetes) [525-527].

Loss-of-function *KCNJ11* mutations associated with PHHI, such as Y12X, are predicted to cause reduced sensitivity to magnesium-ADP or response to a fall in the intracellular ATP/ADP ratio, leaving the channel in a permanently closed/depolarised state and thus causing unregulated insulin secretion [528-530].

It has been postulated that mutations in *KCNJ11* lead to decreased insulin secretion due to reduced sensitivity to ATP and thus gain of channel function [531]. *Kir6.2*[$\Delta N2-30$]mice, which express pancreatic K_{ATP} channels with reduced ATP sensitivity, developed severe early-onset diabetes [529]. Phenotypic variations (neonatal diabetes +/- developmental delay, muscle weakness and epilepsy) are thought to be due to the severity of the reduction in ATP sensitivity [527]. The R201C mutation, identified in individuals with PNDM alone, produced smaller K_{ATP} currents and less change in ATP sensitivity than Q52R and V59G,

associated with 'syndromic PNDM' (PNDM + other features). R201C lies at the putative ATP-binding site and is thought to impair ATP sensitivity directly, whereas Q52R and V59G are thought to impair sensitivity indirectly by increasing the open probability of the channel. Kir6.2 is also found in K_{ATP} channels in skeletal and cardiac muscle and the brain; these channels may only open in situations of metabolic stress, and so a greater reduction in ATP sensitivity may be required to increase the resting K_{ATP} current sufficiently to influence the function of these cells and cause a more severe phenotype [527].

- ii) *Bad* (BCL2-associated agonist of cell death) encodes the minimal death domain BH3-only protein, a member of the B-cell lymphoma 2 protein family [532].

Heterodimerisation of BAD with the antiapoptotic proteins BCL- X_L or BCL-2 at membrane sites prevents them from stopping apoptosis (shown in yeast) [533]. This allows BAX/BAK-triggered apoptosis [534]. An example of this was seen in rats with kainic acid-induced seizures, where dissociation of BAD from its chaperone protein 14-3-3 was induced, allowing it to bind to BCL- X_L and leaving BAX free to translocate to the mitochondria, causing release of cytochrome c from the mitochondria to the cytosol and activation of caspase-9 and caspase-3 [535]. An alternative theory is that BAX/BAK is bypassed, as BAD interacts with BCL- X_L , displacing it from voltage-dependent anion channels, leading to the sensitisation of the mitochondrial permeability transition pore to Ca^{2+} [536]. Serine-phosphorylated BAD (at serine residues Ser112, Ser136 or Ser155) is unable to bind BCL- X_L or BCL-2 [537], leaving BCL-2 free to inhibit BAX-triggered apoptosis [538], promoting cell survival. Phosphorylation of BAD is regulated by various survival factors and death stimuli, such as protein kinase A (PKA), p21-activated protein kinase 1 (PAK1), protein Kinase B (PKB or AKT) and extracellular signal-regulated kinases 1 and 2 (ERK1/2) [539].

Due to its role in regulating apoptosis, potential neuroprotective properties have been assigned to BAD. The antiapoptotic properties of the cytokine tumor necrosis factor-like weak inducer of apoptosis are thought to be due to an induction in the expression of *Bad* phosphorylated at Ser112 in mouse cerebral cortical neurons, mediated by ERK 1/2 [540]. Phosphorylated BAD has also been implicated in the mechanisms of progesterone-mediated protection after ischaemic stroke: the decreased expression of phosphorylated *Bad* in rat cortical penumbra caused by permanent middle cerebral artery occlusion was attenuated by progesterone [541].

BAD also plays a metabolic role through regulation of glucose homeostasis: *Bad*^{3SA/3SA} (non-phosphorylatable) and *Bad*^{-/-} mice exhibited significant fasting hyperglycaemia and a 'marked' defect in blood glucose clearance [542]. BAD has been found to associate with mitochondrial glucokinase in the liver [542, 543] and pancreatic β cells [544] and is essential for glucokinase activity [544]. This activation of glucokinase activity correlates with the ability of BAD to stimulate insulin secretion, a function which is independent from its apoptotic role: *Bad*^{3SA}, whose binding to BCL-2 and BCL-X_L cannot be inhibited, and *Bad* expressing the L151A mutant, which cannot bind to BCL-2 and BCL-X_L, both displayed defective glucose-stimulated insulin secretion [544]. These dual functions of BAD, apoptotic and metabolic, are thought to be dependent on phosphorylation of BAD at Ser155: Ser155 phosphorylation disrupts the interaction between BAD and BCL-X_L [545] but is required for its control of glucose-stimulated insulin-secretion [544]. Furthermore, the fact that the modified BAD BH3 peptide, SAHBA_(L,S,D→A), in which Leu151, Ser155 and Asp156 (required for the metabolic activity of BAD, measured by insulin release) are converted to alanine, was severely compromised in glucokinase binding, shows that the interaction of BAD with glucokinase enables regulation of glucose-stimulated insulin secretion [544].

It is of interest that islet *Bad* mRNA levels were significantly increased in mice fed a high-fat diet, compared to those on a control diet, and that several of the kinases that regulate BAD Ser155 phosphorylation are nutrient sensitive [544]: for example, AKT-p70S6 kinase in β cells is stimulated by glucose and glucagon-like peptide-1, a hormone released in response to feeding that stimulates glucose-associated insulin secretion [544, 546].

4.1.3 Rationale for selection of candidate genes

Bad^{S155A} (non-phosphorylatable) and *Bad*^{-/-} cortical neurons and astrocytes exhibited increased mitochondrial utilisation of BHB and decreased mitochondrial utilisation of glucose, measured by maximal respiration, compared with wild types [226]. These changes in energy substrate utilisation are reminiscent of the response to fasting or KD-/high-fat diet-feeding. The fact that similar changes were seen in *Bad*^{S155A} and *Bad*^{-/-} cell types suggests that this alteration in fuel utilisation is linked to BAD's metabolic, rather than apoptotic role, and may be regulated by its phosphorylation status (*Bad*^{S155A} and *Bad*^{-/-} have opposite effects on BAD's apoptotic function).

Bad^{S155A} and *Bad*^{-/-} mice were resistant to kainic acid- and pentylenetetrazole-induced seizures, compared to wild types. This resistance was diminished in *Bad*^{-/-}, *Kir6.2*^{-/-} mice, indicating that K_{ATP} channels mediate Bad's effect on neuronal excitability and seizure resistance. Consistent with this, the open probability of single K_{ATP} channels was increased in *Bad*-deficient neurons, compared to wild types, and K_{ATP} conductance decreased, compared to wild types or neurons from *Bad*^{-/-}, *Kir6.2*^{-/-} animals. It was postulated that the increase in K_{ATP} channel activity was at least partly due to increased BHB, which links back to the earlier proposed 'K_{ATP}-glycolysis hypothesis' [295], whereby reduced glycolysis (and thus reduced glycolytic ATP production) brought about by low glucose/high ketone body concentrations triggers the opening of K_{ATP} channels (most importantly in the brain), depressing neuronal excitability. It must be noted that there was no significant difference in

total cellular ATP levels between wild type and *Bad*^{-/-} brains [226], although it may be that the local reduction of ATP in the submembrane compartment is more relevant for KD response, rather than levels in system structures or whole organs. K_{ATP} channel activation by other metabolites, such as phosphatidylinositol-4,5-bisphosphate (PIP₂), may also play a role [226, 547]. Whole brain BHB content was shown to be significantly higher in *Bad*^{-/-} mice, compared to controls, indirectly supporting the link between ketone bodies, K_{ATP} channels and seizure protection. Serum BHB was similar in *Bad*^{-/-} and wild type mice and liver-specific knockdown of *Bad* did not provide seizure protection, indicating that local, rather than systemic, changes in ketone body metabolism may be important.

The apoptotic role of *Bad* has also been implicated in the antiepileptic mechanisms of the KD. Due to findings that, following a kainic acid-induced seizure, *Bad* dissociates from 14-3-3 in rat hippocampus [535, 548], expression levels of components in this cell death/survival pathway were measured [549]: although there was no change in hippocampal *Bad* expression levels in KD-fed mice (with or without kainic acid-induced seizure), the KD significantly suppressed the kainic acid-induced reduction in *Bad* phosphorylated at Ser136 (pBAD). The KD was also found to increase interaction between pBAD and 14-3-3, decrease binding of BAD to BCL-X_L, and decrease activation of BAX and caspase-3 – confirmed by decreased levels of cleaved caspase-3, measured by Western blot analysis.

Noh et al. (2006) suggest that, in their study, the *Bad* signalling pathway was modulated by Akt, as the KD attenuated the kainic acid-induced reduction in Akt phosphorylated at Ser473 (which is known to phosphorylate BAD at Ser136), and there were no changes in levels of *Bad* phosphorylated at Ser112 (which may be phosphorylated by ERK1/2 and PKA) between kainic acid-treated KD-fed mice and those fed a standard diet [550]. Consistent with this, the KD significantly reduced *pAkt* expression in rat hippocampus and liver [418]; the authors reported that this inhibition of Akt signalling was a marker of inhibition of the

mTOR pathway. This is supported by findings of increased AMP kinase (which inhibits mTOR via activation of the tuberous sclerosis complex) activity and phosphorylation in the liver and soleus muscle of KD-fed rats [192].

No evidence was found for direct KD-induced effects on the ERK1/2 and PKA pathways, although some related studies may be of interest:

- i) A high-fat diet was shown to have no effect on Erk-1 or -2 expression in mouse white adipose tissue, liver, and muscle, although Erk activity was more than three times higher in white adipose tissue from mice fed a high-fat diet, compared to those fed a standard diet [551]; wild-type mice fed a high-fat diet became glucose intolerant, but *Erk1*^{-/-} mice were protected from this.
- ii) Phosphorylation of Erk1/2 was observed in tumour cells of the adrenal medulla after exposure to butyrate [552].
- iii) *Fgfr1*, whose expression is upregulated by KD-feeding [413], when bound to β Klotho and activated, phosphorylates *Fgfr* substrate 2, which leads to the downstream phosphorylation of Erk1/2 [553].

Independent of the pathway involved, there is supporting evidence that the KD may provide a neuroprotective effect through regulation of the cell death cascade. The KD-induced reduction in rat hippocampal caspase-3 [549] coincides with findings that KD-feeding attenuates an obesity-induced increase in caspase-3 mRNA expression in rat whole brain [415]. The antiapoptotic properties of the KD may also be mediated by downregulation of hippocampal *Pp2ca* expression [290], which encodes an enzyme (PP2A) that may trigger apoptosis by inactivating BCL2 [422]. PP2A has also been shown to catalyse BAD

dephosphorylation (and thus its proapoptotic activity) [539] and has been implicated in the regulation of the phosphorylation status of Akt [554].

Collectively, these findings implicate *KCNJ11* (encoding a subunit that is essential for activation of K_{ATP} channels in the presence of ketone bodies or reduced glycolysis) in the antiepileptic mechanisms of the KD, and introduce the possibility that variation in *BAD* may favourably affect response to the KD, an effect mediated through *KCNJ11*. It remains to be seen whether there is a place for the K_{ATP} /glycolysis hypothesis alongside the antiapoptosis theory, with *BAD* representing a point of convergence for various death/survival pathways. It may be that *BAD* plays a bifunctional (apoptotic vs metabolic [via K_{ATP} channels]) role in the antiepileptic effects of the KD. In this way, *BAD* variation may affect multiple pathways. The possibility also remains that different mechanisms of action come into play in distinct individuals, or in response to diverse situations.

This is the first study to explore the putative association of *KCNJ11* and *BAD* genetic variation with KD response in humans.

4.2 Methods

4.2.1 Phenotypic data

KD response at 3-month follow-up was used as the main phenotype for the candidate gene analysis. As outlined in Chapter 3, using clinic letters and information from patient hospital records, individuals who achieved $\geq 50\%$ seizure reduction were classified as ‘responders’ and those who reported $< 50\%$ seizure reduction as non-responders’. 3-month KD response was chosen because:

- i) All cases received from Harvard Medical School had KD response classified at the 3-month point only.

- ii) The maximum proportion of available data is used - the largest cohort consists of individuals with 3-month response data; this is the case whether Harvard data are used or not.
- iii) In clinical practice, response to the KD is first formally judged at 3-month follow-up.

Another analysis was conducted with summary diet response (*overall* responders and non-responders, taking into account long-term KD response), but data received from Harvard could not be used for this.

Criteria for the summary diet response phenotype are outlined below:

Responders:

- Cases who achieved $\geq 50\%$ seizure reduction at every time point recorded

Non-responders:

- Cases who achieved $< 50\%$ seizure reduction at every time point recorded

Excluded:

- Cases who consistently reported approximately 50% seizure reduction over time, as responder/non-responder status is ambiguous
- Cases with fluctuating diet response over time

Extreme KD response was not included as a phenotype, as this association analysis is concerned with the effects of common variants in *KCNJ11* and *BAD*, which are not expected to have a large phenotypic effect: purifying selection acts against deleterious variants [555] and, of all recently arisen variants, only the frequency of those with the most deleterious effect could have been affected by selection.

4.2.2 Genetic data

DNA samples were diluted to 40µL at 50ng/µL by the researcher and sent to the Royal Devon & Exeter Hospital for sequencing of *KCNJ11* and *BAD*, in two batches. Exon 1 of *KCNJ11* was amplified in three fragments, and the three exons were amplified in *BAD* using M13 tailed PCR primers. Sequencing was performed using a Big Dye Terminator Cyclor Sequencing Kit (Applied Biosystems, Warrington, UK) according to manufacturer's instructions. Reactions were analysed on an ABI 3130 Capillary sequencer (Applied Biosystems, Warrington, UK) and sequences were compared to the published sequence (*KCNJ11*: NM_000525; *BAD*: NM_004322.3) using Mutation Surveyor v.4.0.6 (SoftGenetics, Pennsylvania, US). Identified variants were checked against known polymorphisms and mutations [556] and variant databases (dbSNP, the 1000 Genomes Project and Exome Variant Server).

KCNJ11 sequencing results were also received from Harvard Medical School for a further 58 cases. The full *KCNJ11* gene had been sequenced bi-directionally using three 500-600bp amplicons with published primers [557]. Sequencing was performed using an ABI 377 machine (PE Biosystems, NYSE:PEB) with Big Dye terminator chemistries (PE Biosystems, NYSE:PEB). Sequence traces were analysed using the GAP4 program of the Staden package (<http://www.mrc-lmb.cam.ac.uk/pubseq>).

The total number of cases with candidate gene sequencing and diet response data is outlined in Figure 4.1.

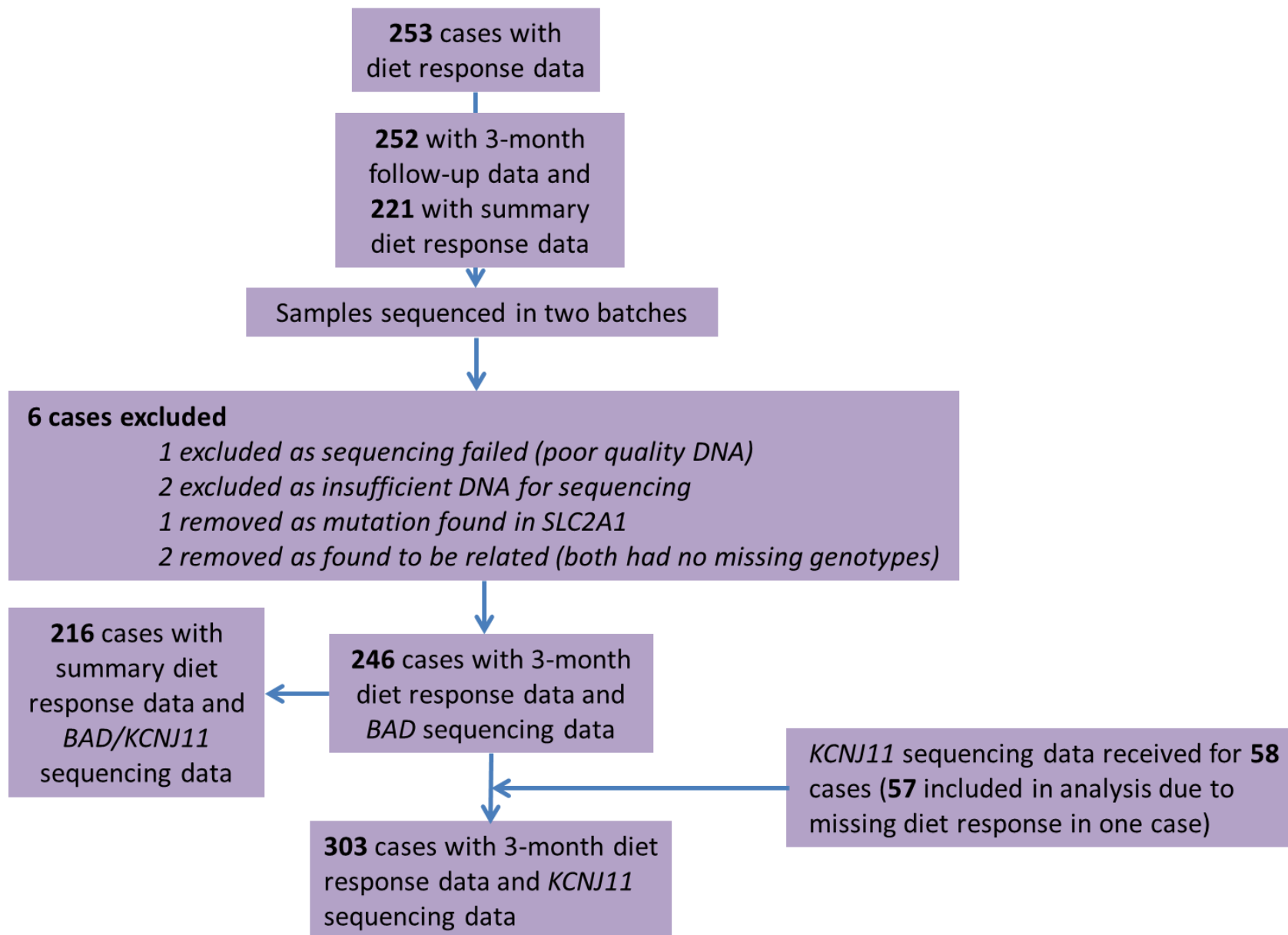


Figure 4.1: Flow-chart of cases with diet response data and included in final candidate gene analyses

4.2.2.1 Quality control

Relevant per-SNP and per-individual quality control measures, as outlined in [510], were conducted in PLINK (v1.07, <http://pngu.mgh.harvard.edu/purcell/plink>, [558]):

SNPs with deviation from Hardy Weinberg equilibrium (HWE), $p\text{-value} < 1 \times 10^{-6}$, were excluded (PLINK command: `--hwe-all --hwe 0.000001`). If not in HWE, variant allele frequencies are not expected to remain constant over time and the population is not behaving like a single randomly-mating unit without intense viability selection acting on the sampled loci [559]).

Duplicate and related persons were excluded. Identity-by-descent (IBD) was calculated (PLINK command: `--genome`) between all pairs of individuals with genome-wide SNP data (see Chapter 5) and cases with Pi-hat [proportion of IBD] score > 0.4 (duplicated cases and full sibling pairs) were excluded. All cases recruited by the researcher had genome-wide SNP data available, with the exception of three participants, for whom blood samples were received after the last batch was sent to AROS for genotyping with the array. Cases received from Harvard did not have genome-wide SNP data available.

SNPs or individuals with $> 2\%$ missing rates were excluded (PLINK command: `--mind 0.02 --geno 0.02`); Harvard cases were analysed separately from other participants for this step, as they did not have *BAD* sequencing data.

4.2.2.2 Association analysis

For those SNPs with $MAF > 0.01$ in this cohort, a Fisher's allelic association exact test was performed in PLINK using the `--fisher` command, to determine whether allele frequency differed between responders and non-responders. This test is suited to small sample sizes

and avoids bias due to distributional approximation. The MAF cut-off point of 0.01 allowed SNPs of common and intermediate frequency in this cohort to be included in the association analysis. SNPs present in <1% of the cohort were not included in the analysis, as the variant would have to have a large effect size in order for the study to be appropriately powered to detect an association [560]; also, any association detected would have been driven by the genotypes of only a few individuals and would therefore be less robust [510].

Allelic odds ratios were calculated (command: *--ci 0.95*), which compare the odds of favourable or unfavourable KD response in an individual carrying the minor allele to the odds in an individual carrying the major allele [561]. A 95% confidence interval (CI) was applied to indicate the reliability of the odds ratio estimate, as is widely used in biostatistics, indicating the probability that the values within the CI capture the true population parameter.

P-values were obtained from 100,000 permutations, using *the --mperm 100000* command. Permutation tests are considered the gold-standard when adjusting for multiple testing in genetic association studies [562]. Phenotypes are randomly shuffled with respect to genotypes among observed data, removing any genotypic-phenotypic association; each random shuffle of the data represents one possible sampling of individuals under the null hypothesis of no association, a process which is repeated a certain number of times [464, 563]. The higher the number of permutations, the greater the precision of the generated p-value, although this has been debated [564]. ≥10,000 permutations are needed to estimate a 0.01 threshold [565, 566]. In recent candidate gene studies, the number of permutations varies, ranging from 1,000 [567] to 100,000 [568] to 1,000,000 [569]. Here, 100,000 permutations have been chosen, in order to obtain as precise p-values as possible whilst remaining realistic with regards to time needed for these computationally-intensive

permutation procedures (computation was not overly intensive in this case, due to the small number of SNPs to be tested).

PLINK was used to conduct a haplotype analysis, including all previously known SNPs. No MAF threshold was applied because haplotype-based methods have higher power to detect an association between rare variants and a phenotypic trait, compared to single-SNP methods [570, 571]. Haplotype analyses may capture epistatic interactions between SNPs at a locus [508] and have higher power to detect an association than single SNP analyses in the presence of multiple susceptibility alleles, although this advantage may be lost when there is high linkage disequilibrium (LD) between SNPs [507]. The *--hap-assoc* command was used, which compares allele frequencies between cases and controls, along with the *--hap-window* command, which specifies all haplotypes in sliding windows across the dataset. Permutation of p-values cannot be conducted with this test in PLINK and so a Bonferroni-adjusted threshold was set, which accounts for the number of independent SNPs included in the haplotype analysis ($n=13$), with an alpha value of 0.05 ($0.05/13=0.0038$).

4.2.2.3 Population structure

The 1000 Genomes Project browser (<http://browser.1000genomes.org>) was used to investigate population genetics for variants included in the Fisher's test. Variants that displayed statistically significant differences in population-specific allele frequencies (shown by chi-squared test conducted in Microsoft Excel 2010 (v. 14, Microsoft, Washington, USA) were analysed with a Cochran-Mantel-Haenszel test, using the *--mh --within* command in PLINK. This is a single-SNP stratified association analysis, which tests for overall disease/gene association in case/control studies, conditional on strata/clusters (in this case, self-reported ethnicity), and is suitable both for a large number of small clusters

and a small number of large clusters. It has been found to adequately control for population stratification for variants of different MAFs, used as a reference to assess performances of other stratified association tests, as it adjusts exactly on the confounding factor introduced in the analysis [572]. This approach has been used in Wellcome Trust Case Control Consortium datasets to adjust for population substructure [573].

The self-reported ethnicity categories in this cohort were:

- 1 - Caucasian (including Australian and American Caucasians), n=251
- 2 – African, n=5
- 3 - Middle Eastern, n=6
- 4 - Central/South Asian, n=14
- 5 - East Asian, n=2
- 6 - Black and Caucasian mix, n=19
- 7 - East Asian and Caucasian mix, n=3
- 8 - South Asian and Caucasian, n=2
- 9 - South American, n=1

The Cochran-Mantel-Haenszel test cannot be used for haplotype analyses in PLINK.

4.2.2.4 Information on variants found

R^2 values for all possible pairs of previously reported SNPs present in this cohort were calculated using PLINK, using the `--r2 --ld-snp-list` command. R^2 values, which range between 0 and 1, represent a statistical measure of LD correlation between variants [464] - a higher r^2 value indicates a greater degree of LD between two markers.

Predictions of the functional impact of previously reported variants found in *KCNJ11* and *BAD* in this cohort were obtained from various algorithms using WANNONAR (<http://wannovar.usc.edu>, [574]): SIFT (Sorting Intolerant From Tolerant) [575], PolyPhen2

(Polymorphism Phenotyping v2) [576], LRT (likelihood ratio test) [577], Mutation Taster [578] and GERP++ (Genomic Evolutionary Rate Profiling) [579]. These scores were only available for non-synonymous variants. GERP++ scores for synonymous variants were obtained from National Heart, Lung, and Blood Institute (NHLBI) Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP) (<http://evs.gs.washington.edu/EVS>). Variants were also looked up in Phenotype-Genotype Integrator (<http://www.ncbi.nlm.nih.gov/gap/phegeni>) to obtain information regarding human phenotypic-genotypic associations and function class of the variants, and in dbSNP (<http://www.ncbi.nlm.nih.gov/SNP/index.html>) to obtain MAF information.

The predicted functional impact of previously unreported *KCNJ11* and *BAD* variants, in terms of function class, SIFT score, nucleotide and amino acid conservation, and Mutation Taster score, was provided by Alamut (v2.2, Interactive Biosoftware LLC, Rouen, France). The aim of this was to aid interpretation of results, prioritising SNPs with a higher predicted impact on gene and protein function and, theoretically, the phenotypic trait in question [580].

4.2.3 Power calculations

Power calculations for the 3-month KD response cohort were conducted using PGA Power Calculator [581]. Curves depicting potential detectable relative risk and MAF in this cohort, with 80% power, were plotted. Power sample size curves using the MAF of the most significant SNP in the association analysis were also plotted. An alpha level according to the effective number of tests was used (0.05/7, as seven SNPs that were not in high LD were tested for in the association analyses).

As there is no naturally-existing KD (or indeed sustained calorie restriction, at least in Europeans), disease prevalence (or prevalence of the trait) is hard to gauge. The disease prevalence of epilepsy (0.5%) was therefore used, as in other studies [582].

The inheritance model is unknown and so power calculations were performed using co-dominant (where the two alleles have an equal effect on the phenotype), dominant (where one allele masks the effect of the other allele on the phenotype) and recessive (where the phenotype is only seen if the individual is homozygous for the risk allele) penetrance models.

4.3 Results

4.3.1 Quality control

No SNPs were removed due to deviation from HWE or missingness rates. Two cases were removed due to relatedness (both had no missing genotypes - in the case of different missingness rates, the subject with the lower rate would have been excluded). One subject was removed due to presence of *SLC2A1* mutation (identified from whole exome sequencing, after recruitment and candidate gene sequencing). *KCNJ11* and *BAD* sequencing failed in two subjects.

4.3.2 Common variation in *KCNJ11* and *BAD*

4.3.2.1 Overview

In this cohort, nine previously-reported SNPs were found in *KCNJ11* and five in *BAD*. These are given in Table 4.1, along with their predicted functional impact.

Variants were also looked up in The Human Gene Mutation Database (<http://www.hgmd.cf.ac.uk>, [583]) and NCBI (National Center for Biotechnology Information) Variation Viewer (www.ncbi.nlm.nih.gov/sites/varvu). rs5219, rs5215 and rs5218 were all reported to be associated with Type II diabetes in The Human Gene Mutation Database. rs5219 was classified as a 'drug-response' variant in NCBI Variation Viewer, rs5215 had no clinical interpretation (although 21 publications were linked to the

variant, which analysed its association with Type II diabetes) and rs5218 was not present in NCBI Variation Viewer.

Despite being classified as disease-causing by MutationTaster, rs1800467 was not present in either database, and rs41282930 was classified as 'probable non-pathogenic' in NCBI Variation Viewer and was not present in The Human Gene Mutation Database.

Table 4.1: Previously reported SNPs found in KCNJ11 and BAD in Ketogenic diet cohort

Gene	SNP rs number	Location (build 37/hg19)	Variant nomenclature (cDNA level and protein level)*	Function class	Minor Allele Frequency (1000 Genomes Project)	dbNSFP SIFT score and class	Polyphen2 score and class	LRT score and class	Conservation score GERP++	Mutation Taster score and class
KCNJ11	rs5219	11: 17,409,572	NM_000525.3:c.6 7A>G p.Glu23Lys	missense	0.274	0.75 Tolerated	0 Benign	0.927728 Neutral	3.65	0 Polymorphism
	rs5218	11: 17,409,069	NM_000525.3:c.5 70C>T p.Ala190Ala	synonymous	0.275	n/a	n/a	n/a	-0.1	n/a
	rs5216	11: 17,408,838	NM_000525.3:c.8 01C>G p.Leu267Leu	synonymous	0.017	n/a	n/a	n/a	2.21	n/a
	rs1800467	11: 17,408,831	NM_000525.3:c.8 08C>G p.Leu270Val	missense	0.022	0.98 Deleterious	0 Benign	0.999656 Deleterious	4.79	0.687162 Disease- causing
	rs5215	11: 17,408,630	NM_000525.3:c.1 009G>A p.Val337Ile	missense	0.280	0.76 Tolerated	0.002 Benign	0.947433 Neutral	-2	1.00x10 ⁻⁵ Polymorphism -automatic
	rs8175351	11: 17,408,496	NM_000525.3:c.1 143G>A p.Lys381Lys	synonymous	0.006	n/a	n/a	n/a	4.26	n/a
	rs4128293 0	11: 17,408,485	NM_000525.3:c.1 154C>G p.Ser385Cys	missense	0.0114	1 Deleterious	0.27 Benign	0.696141 Neutral	4.52	0.630056 Disease- causing
	rs5214	11: 17,408,550	NM_000525.3:c.1 089A>G p.Ser363Ser	synonymous	0.003	n/a	n/a	n/a	-10.4	n/a
	rs1120704 96	11: 17,409,531	NM_000525.3:c.1 08G>A p.Val36Val	synonymous	0.002	n/a	n/a	n/a	3.66	n/a

BAD	rs3488200 6	11: 64,051,823	NM_004322.3:c.1 8G>A p. = (p.Glu6Glu)	synonymous	0.0133	n/a	n/a	n/a	4.79	n/a
	rs7505260 0	11: 64,051,853	NM_004322.3:c.- 13C>T p.?	5' UTR	0.0137	not found	not found	not found	0.8	not found
	rs2286615	11: 64,039,175	NM_004322.3:c.2 88C>A p.= (p.Arg96Arg)	synonymous	0.0934	n/a	n/a	n/a	3.87	n/a
	rs2286616	11: 64,039,136	NM_004322.3:c.3 27C>A p.= (p.Arg109Arg)	synonymous	0.006	n/a	n/a	n/a	-2.44	n/a
	rs5928362 9	11: 64,037,679	NM_004322.3:c.* 2C>T p.?	3' UTR	0.0201	not found	not found	not found	not found	not found

*according to Human Genome Variation Society recommendations for the description of DNA sequence variants - v2.0 <http://www.hgvs.org/mutnomen/recs-DNA.html>

SIFT Pathogenicity score from dbNSFP: tolerated < 0.95, deleterious > 0.95

PolyPhen2 Pathogenicity score from dbNSFP: probably damaging > 0.85, possibly damaging 0.85-0.15, benign < 0.15

GERP++ Nucleotide conservation score from dbNSFP GERP: higher number is more conserved, > 0 is generally conserved

LRT Pathogenicity probability score from dbNSFP: closer to 1 is more likely to be damaging. Deleterious fulfils the following criteria: (i) from a codon defined by LRT as significantly constrained (original LRT score <0.001 and <1), (ii) the AA position has more than 10 eutherian mammal alignments. Neutral: the alternative AA is presented in at least one of the eutherian mammals, or the codon is not defined by LRT as significantly constrained; otherwise unknown.

MutationTaster Pathogenicity probability score from dbNSFP: closer to 1 is more likely to be damaging

Due to its function class, SIFT, PolyPhen2, LRT and Mutation Taster classifications, and location in a conserved region, rs1800467 in *KCNJ11* is considered most likely to have a deleterious effect on gene function, although it is not present in The Human Gene Mutation Database. rs41282930 in *KCNJ11* is of interest, although two out of the five functionality prediction algorithms used have classified this variant as 'benign' or 'neutral'. Due to their association with Type II diabetes, rs5219, rs5215 and rs5218 are also of interest.

The details displayed in Table 4.1 will aid interpretation of association results; if an association is reported, the potential effect of the variant on the gene or its expression may be judged. No variants will be excluded from the association analysis based on function class, predicted deleterious, or presence in mutation databases.

As shown in Table 4.2, only rs5219 and rs5215 are in high LD in this cohort. One of these SNPs could be removed from the association analysis, as knowing the alleles present at one SNP provides knowledge of those at the other SNP.

As most SNPs are not in high LD with each other, this indicates that the predictive power of haplotypes, as well as single SNP tests, may be exploited.

Table 4.2: R² values between pairs of previously reported *KCNJ11* and *BAD* SNPs found in Ketogenic diet cohort

<i>KCNJ11</i> SNPs											<i>BAD</i> SNPs				
<i>KCNJ11</i> SNPs		rs112070496	rs41282930	rs8175351	rs1800467	rs5219	rs5218	rs5216	rs5215	rs5214	rs75052600	rs59283629	rs34882006	rs2286616	rs2286615
	rs112070496	1	0.0001	0.0002	0.0003	0.0031	0.0024	0.0002	0.0030	1	0.0000	0.0000	0.0003	0.0000	0.0015
	rs41282930	-	1	0.0224	0.0086	0.0012	0.0006	0.0276	0.0012	0.0001	0.0000	0.0001	0.0007	0.0001	0.0169
	rs8175351	-	-	1	0.0000	0.0082	0.0093	0.0027	0.0083	0.0002	0.0001	0.0177	0.0027	0.0002	0.0006
	rs1800467	-	-	-	1	0.0519	0.0112	0.0000	0.0507	0.0003	0.0003	0.0009	0.0019	0.0006	0.0028
	rs5219	-	-	-	-	1	0.1657	0.0259	0.9933	0.0031	0.0036	0.0000	0.0006	0.0019	0.0083
	rs5218	-	-	-	-	-	1	0.0018	0.1613	0.0024	0.0020	0.0004	0.0111	0.0062	0.0031
	rs5216	-	-	-	-	-	-	1	0.0260	0.0002	0.0002	0.0006	0.0001	0.0004	0.0044
	rs5215	-	-	-	-	-	-	-	1	0.0030	0.0036	0.0000	0.0005	0.0019	0.0076
	rs5214	-	-	-	-	-	-	-	-	1	0.0000	0.0000	0.0003	0.0000	0.0015
<i>BAD</i> SNPs	rs75052600	-	-	-	-	-	-	-	-	-	1	0.0000	0.0003	0.0000	0.0015
	rs59283629	-	-	-	-	-	-	-	-	-	-	1	0.0010	0.0001	0.0045
	rs34882006	-	-	-	-	-	-	-	-	-	-	-	1	0.0007	0.0013
	rs2286616	-	-	-	-	-	-	-	-	-	-	-	-	1	0.0030
	rs2286615	-	-	-	-	-	-	-	-	-	-	-	-	-	1

4.3.2.2 Population genetics

Population genetics data from the 1000 Genomes Project were available for all variants used in the association analysis, with the exception of rs5219 (although population genetics information for this SNP may be inferred from rs5215). As shown in Table 4.3, all SNPs have significant population-specific differences (according to the 1000 Genomes Project) in allele frequencies.

Table 4.3: Observed and expected population genetics data for *KCNJ11* and *BAD* SNPs (from the 1000 Genomes Project) and results from Chi-square test of independence

Gene	SNP rs number	Population	Observed counts for allele 1	Observed counts for allele 2	Expected counts for allele 1*	Expected counts for allele 2**	P-value from Chi-squared distribution
<i>KCNJ11</i>	rs5218	African	32	952	271	713	6.9×10^{-120}
		Asian	532	612	315	829	
		American	132	592	199	525	
		European	506	1010	417	1099	
	rs5216	African	4	980	17	967	1.66×10^{-12}
		Asian	20	704	12	712	
		American	0	1144	19	1125	
		European	50	1466	26	1490	
	rs1800467	African	2	982	22	962	1.52×10^{-18}
		Asian	0	1144	25	1119	
		American	28	696	16	708	
		European	66	1450	33	1483	
	rs5215	African	54	930	275	709	2.36×10^{-71}
		Asian	434	710	320	824	
		American	224	500	203	521	
		European	510	1006	424	1092	
	rs817351	African	978	6	978	6	0.0027
		Asian	1144	0	1137	7	
		American	714	10	719	5	
		European	1504	12	1506	10	
		Asian	0	1144	13	1131	
		American	12	712	8	716	
		European	6	1510	17	1499	
		European	6	1510	17	1499	
<i>BAD</i>	rs34882006	African	492	0	485	7	6.78×10^{-7}
		Asian	572	0	564	8	
		American	356	6	357	5	
		European	735	23	748	10	
	rs2286615	African	11	482	46	447	1.22×10^{-15}
		Asian	40	532	53	519	

	American	32	330	34	328
	European	121	637	71	687

*expected counts for allele 1 were calculated by the sum of observed counts of alleles 1 and 2 in the specific population, multiplied by the total count for allele 1 (in all populations), all divided by the total observed counts of alleles 1 and 2 (in all populations)

**expected counts for allele 2 were calculated by the sum of observed counts of alleles 1 and 2 in the specific population, multiplied by the total count for allele 2 (in all populations), all divided by the total observed counts of alleles 1 and 2 (in all populations)

As shown in Table 4.4, the number of responders/non-responders observed in each population in this cohort is not significantly different ($p\text{-value} > 0.5$) to expected numbers, although there were only small numbers of individuals in most groups. When comparing observed and expected numbers of responders/non-responders in all population groups simultaneously, $p\text{-value} = 0.66$, demonstrating that there is no significant difference in responder/non-responder rates among different populations in this cohort.

However, due to potential population-specific allele frequencies, stratified association analyses were also conducted with self-reported ethnicity as strata.

Table 4.4: Breakdown of Ketogenic diet responders and non-responders by ethnicity, n=303

Ethnicity	Observed number of responders	Observed number of non-responders	Expected number of responders*	Expected number of non-responders**	P-value from Chi-squared distribution
Caucasian	148	103	146	105	0.78
Black African	2	3	3	2	0.41
Middle Eastern	3	3	3	3	0.69
Central/South Asian	7	7	8	6	0.54
East Asian	2	0	1	1	0.23
Black African and Caucasian mix	10	9	11	8	0.63
East Asian and Caucasian mix	2	1	2	1	0.76
South Asian and Caucasian mix	2	0	1	1	0.23

South American	0	1	1	0	0.24
-----------------------	---	---	---	---	------

*expected number of responders were calculated by the sum of observed number of responders and non-responders in the specific ethnic group, multiplied by total number of responders (in all ethnic groups), all divided by the total number of responders and non-responders (in all ethnic groups)

**expected number of non-responders were calculated by the sum of observed number of responders and non-responders in the specific ethnic group, multiplied by total number of non-responders (in all ethnic groups), all divided by the total number of responders and non-responders (in all ethnic groups).

4.3.3 Association analysis

Results of the association analysis, including SNPs with $MAF > 0.01$, are given in Table 4.5.

The lowest p-value was for rs5216: the G allele was more frequently present in KD non-responders than responders. 16 individuals were heterozygous for the minor allele (0 were homozygous), 2 of which were non-responders and 14 were responders. The unadjusted p-value was < 0.05 , but the p-value obtained from 100,000 permutations was not statistically significant.

Table 4.5: Results of association analysis: common and intermediate variation in *KCNJ11* and *BAD* in Ketogenic diet responders and non-responders (n=303 for *KCNJ11* and n=246 for *BAD*)

Gene	SNP rs number	Location (build 37/hg19)	Minor allele (in Ketogenic diet cohort)	Frequency of minor allele in non-responders	Frequency of minor allele in responders	Unadjusted p-value	Odds ratio [95%CI]	P-value (100,000 permutations)
<i>KCNJ11</i>	rs8175351	11: 17,408,496	A	0.02756	0.02557	1	1.08 [0.3969-2.939]	1
	rs1800467	11: 17,408,831	G	0.07087	0.04261	0.1481	1.714 [0.8466-3.468]	0.6355
	rs5219 (same as for rs5215)	11: 17,409,572	A	0.3268	0.3381	0.7938	0.9504 [0.6745-1.339]	0.9997
	rs5218	11: 17,409,069	T	0.248	0.304	0.1429	0.7552 [0.5247-1.087]	0.6122
	rs5216	11: 17,408,838	G	0.007874	0.03977	0.01872	0.1916 [0.04316-0.8507]	0.132
<i>BAD</i>	rs34882006	11: 64,051,823	A	0.05	0.03175	0.3645	1.605 [0.6445-3.998]	0.9406
	rs2286615	11: 64,039,175	T	0.1125	0.1706	0.07133	0.6161 [0.3671-1.034]	0.362

All SNPs were analysed including ethnicity as a covariate. As shown in Table 4.6, p-values were similar to those obtained when ethnicity was not considered.

Table 4.6: Results of association analysis: common and intermediate variation in *KCNJ11* and *BAD* in Ketogenic diet responders and non-responders, including ethnicity as a covariate (n=303 for *KCNJ11* and n=246 for *BAD*)

Gene	SNP rs number	Unadjusted p-value	Odds ratio [95%CI]	P-value (100,000 permutations)
<i>KCNJ11</i>	rs8175351	0.9842	1.01 [0.369-2.766]	1
	rs1800467	0.1668	1.656 [0.8072-3.396]	0.7343
	rs5219 and rs5215	0.9602	1.009 [0.7099-1.434]	1
	rs5218	0.1712	0.7728 [0.5342-1.118]	0.7514
	rs5216	0.01847	0.1975 [0.0444-0.8784]	0.1318
<i>BAD</i>	rs34882006	0.3133	1.598 [0.6387-3.997]	0.9329
	rs2286615	0.07559	0.6219 [0.3682-1.05]	0.4272

No SNPs were significantly associated with summary KD response, as shown in Table 4.9 and Table 4.10 in Appendix 4.1 and 4.2. rs8175351 had the lowest p-value.

4.3.4 Previously unreported variants or SNPs with minor allele frequency <0.01

KD response in cases with previously unreported variants or those with MAF<0.01 in this cohort, along with the predicted functional impact of these variants, is given in Table 4.7.

NM_00525.3:c.[451G>A];[=], p.Val151Met is of most interest, as it encodes a highly-conserved amino acid (the amino acid is conserved up to Tetraodon), has a deleterious SIFT classification and is classified as disease-causing by Mutation Taster. This variant was found in one non-responder. The missense variant, rs41282930, is also of interest due to its predicted high functional impact, and was present in individuals with a range of responses to the KD.

Independent of predicted functional impact, there does not seem to be a consistent effect of these novel/rare variants on KD response singly, although the cohort size is too small to assess significance.

Table 4.7: KD response in cases with previously unreported variants or variants with MAF<0.01 (in this cohort) in *KCNJ11* and *BAD*

Gene	Variant nomenclature (cDNA level and protein level)* or rs number (dbSNP)	Location build	Predicted function class	SIFT pathogenicity score	Conserved?*	Mutation Taster classification	Number of cases with variant (genotype)	Description of diet response
KCNJ11	NM_00525.3:c.[1116C>T];[=] p.= (p.Ser372Ser)	11:17408523	synonymous	-	-	-	1 (CT)	Case 1: No change in seizures. Weaned off diet after 6 months
	NM_00525.3:c.[912C>T];[=] p.= (p.Tyr304Tyr)	11:17408727	synonymous	-	-	-	1 (CT)	Case 2: No change in seizures. Weaned off diet after 3 months
	NM_00525.3:c.[451G>A];[=] p.Val151Met	11:17409188	missense	Deleterious (score 0.00)	Weakly conserved nucleotide; highly conserved amino acid	Disease-causing (p-value 0.991)	1 (GA)	Case 3: No change in seizures. Weaned off diet after 3 months
	rs5214	11:17408550	synonymous	-	-	-	1 (GA)	Case 4: Partial responder in first 3 months, then improved response in second 3 months (>75% seizure reduction). Currently 6 months on diet.
	NM_00525.3:c.[291C>T];[=] p.= (p.His97His)	11:17409348	synonymous	n/a	n/a	n/a	1 (CT)	Case 5: Around 40% reduction in seizures, and much more alert and interactive. Currently on diet for 3 months
	rs112070496	11:17409531	synonymous	-	-	-	1 (AG)	Case 6: Partial responder in first 3 months, then improved response in second 3 months (>75% seizure reduction). Currently 6 months on diet. (same case as with rs5214)
	NM_00525.3:c.[817A>G];[=] p.Ser273Gly	11:17408822	missense	Deleterious (score 0.03)	Not conserved nucleotide; Highly conserved amino acid	Polymorphism (p-value 1.0)	1 (AG)	Case 7: Partial responder at the start, and more alert, interactive and vocal. Slight reduction in seizures did not last, so weaned off diet after 6 months

	rs41282930	11:17408485	missense	Deleterious (score 0.00)	GERP++ score 4.52	Disease-causing	6 (all GC)	<p>Case 8: According to seizure diary, seizures increased by 386% in first 3 months, compared to baseline. Parents describe that the diet had had no effect on seizure frequency, but the patient had increased awareness</p> <p>Case 9: >50% seizure reduction in first 3 months. Seizure frequency increased during following 3 months and the patient was weaned off diet. Seizure frequency and severity further increased. Patient went back on diet but no improvement.</p> <p>Case 10: Achieved >50% seizure reduction with KD (summary of response over time)</p> <p>Case 11: Achieved >90% seizure reduction with KD (summary of response over time)</p> <p>Case 12: Achieved >50% seizure reduction with KD (summary of response over time)</p> <p>Case 13: Achieved >50% seizure reduction with KD (summary of response over time)</p>
BAD	NM_004322.3:c.[142G>A];[=] p.Ala48Thr	11:64051699	missense	Tolerated (score 0.68, median 4.32)	Not conserved nucleotide; moderately conserved amino acid	Polymorphism (p-value 0.997)	1 (GA)	Case 14: No change in seizures (possible increase). Weaned off diet after 3 months
	rs75052600	11:64051853	5' UTR	unknown	unknown	unknown	1 (CT)	Case 15: No change in seizures, tired and irritable. Weaned off diet before 3-month point
	NM_004322.3:c.-41G>A;[=] p.?	11:64051881	5'UTR	unknown	unknown	unknown	1 (GA)	Case 16: Responder (50-75% seizure reduction). Weaned off diet after 6

								months due to abdominal pain but went back on diet – initial good response
	rs2286616	11:64039136	synonymous	-	-	-	2 (Both CT)	Case 17: Extreme responder – almost seizure-free. Seizure frequency variable after 3 years on diet Case 18: Slight improvement in seizures but weaned off diet after 3 months due to compliance difficulties
	NM_004322.3:c.[456G>A];[=] p.= (<i>p.Gln152Gln</i>)	11:64037732	synonymous	-	-	-	1 (GA)	Case 19: Responder (50-75% seizure reduction). Remains on diet long-term.
	rs59283629	11:64037679	3' UTR	unknown	unknown	unknown	3 (All CT)	Case 20: Responder (>75% seizure reduction). One year on diet Case 21: No change in seizures. Currently at the 3-month point and weaning off diet Case 22: No change in seizures, but much more alert. Currently on diet for 6 months
	NM_004322.3:c.[226T>A];[=] p.Tyr76Asn	11:64039237	missense	Deleterious (score 0.00, median 4.32)	Weakly conserved nucleotide; highly conserved amino acid	Polymorphism (p-value 0.537)	2 (Both TA)	Case 23: Seizure-free. Currently 6 months on diet Case 24: No change in seizures. Off diet after 3 months

*according to Human Genome Variation Society recommendations for the description of DNA sequence variants - v2.0 <http://www.hgvs.org/mutnomen/recs-DNA.html>

** Predictions of conservation are provided by Alamut for novel variants, based on PhyloP score (nucleotide conservation) and cross-species alignment (amino acid conservation)

SIFT Pathogenicity score from wANNOVAR: the closer the number is to 0, the more damaging the variant

PolyPhen2 Pathogenicity score from wANNOVAR: probably damaging >0.85, possibly damaging 0.85-0.15, benign <0.15

GERP++ Nucleotide conservation score from wANNOVAR: higher number is more conserved, >0 is generally conserved

Mutation Taster p-value is an indication of the probability of the prediction. A value close to 1 indicates a high 'security' of the prediction

<http://www.mutationtaster.org/info/documentation.html>. P-values not provided in wANNOVAR.

4.3.5 Haplotype analyses

Haplotype tests yielded similar p-values to the single SNP tests, none of which were below the Bonferroni-adjusted threshold of 0.0038. The haplotypes with the 10 lowest p-values, as shown in Table 4.8, all contained rs5216, and the haplotype with the lowest p-value was rs5216|rs5215|rs5214 (GAA). Apart from the single SNP haplotype (only containing rs5216), all haplotypes also contained rs5215.

Table 4.8: Top 10 most significant results for haplotype association analysis: variation in *KCNJ11* and *BAD* in Ketogenic diet responders and non-responders (n=303 for *KCNJ11* and n=248 for *BAD*)

SNPs in haplotype	Haplotype	Frequency in non-responders	Frequency in responders	Unadjusted p-value
rs5216 rs5215 rs5214	GAA	0.007874	0.03989	0.01544
rs5216	G	0.007874	0.03977	0.01566
rs5216 rs5215	GA	0.007874	0.03977	0.01566
rs5216 rs5215 rs5214 rs75052600	GAAC	0.007905	0.03989	0.01575
rs5216 rs5215 rs5214 rs75052600 rs59283629	GAACC	0.007969	0.04004	0.01608
rs5218 rs5216 rs5215 rs5214 rs75052600 rs59283629	CCAACC	0.4216	0.3312	0.02384
rs5219 rs5218 rs5216 rs5215 rs5214 rs75052600 rs59283629	GCCAACC	0.4217	0.3315	0.02416
rs5218 rs5216 rs5215 rs5214 rs75052600	CCAAC	0.419	0.3305	0.0265
rs5219 rs5218 rs5216 rs5215 rs5214 rs75052600	GCCAAC	0.419	0.3305	0.0265
rs41282930 rs8175351 rs1800467 rs5219 rs5218 rs5216 rs5215 rs5214 rs75052600 rs59283629	CGCGCCAACC	0.3852	0.2989	0.02864

In the summary KD response haplotype analysis, no significant results were obtained (see Table 4.11 in Appendix 4.3).

4.3.6 Power calculations

As shown by Figure 4.2, with a sample size of 303 (as used in the 3-month KD response-*KCNJ11* analysis), variants with a MAF of 0.075 and a relative risk of approximately 2.5 could be detected with 80% power; assuming a dominant penetrance model, variants with

a MAF of approximately 0.1 with a relative risk of approximately 2.5 could be detected; assuming a recessive penetrance model, the MAF of the variant would have to be at least 0.34-0.4 if the variant had the same relative risk.

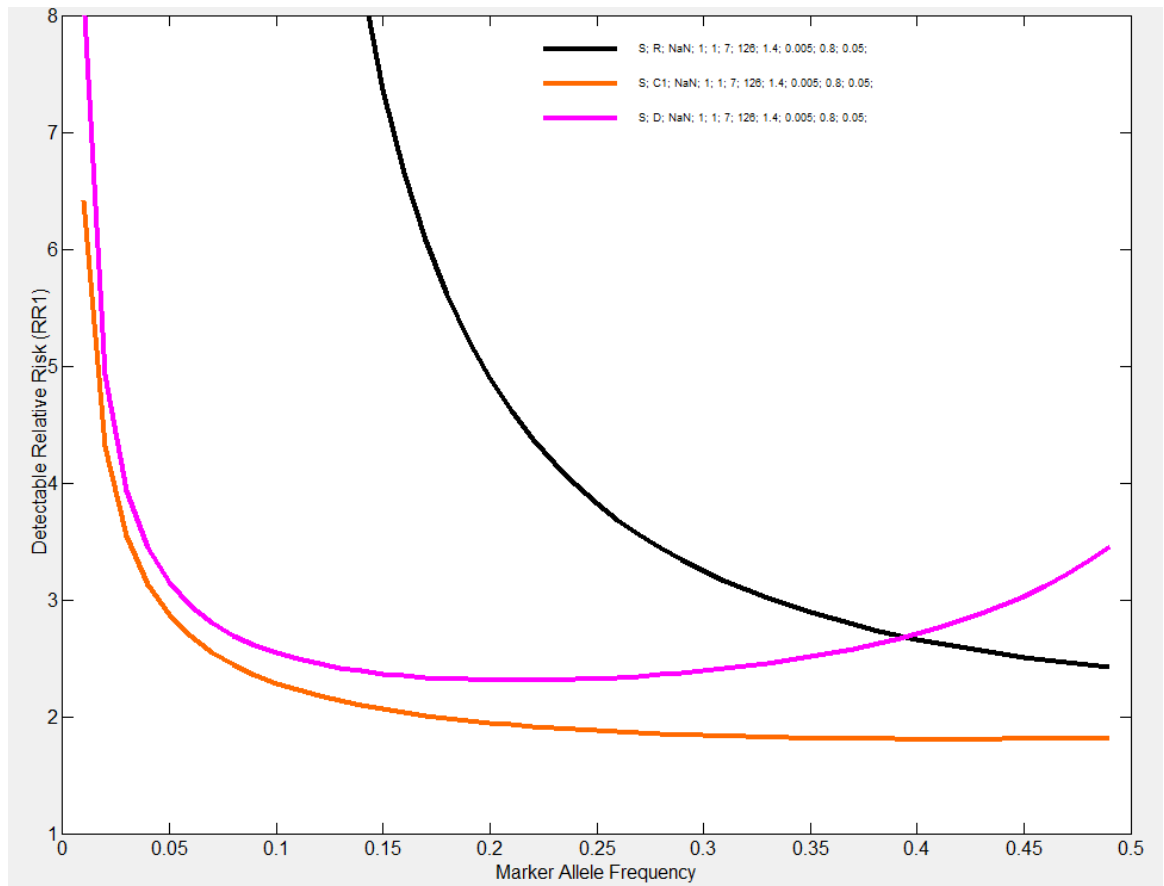


Figure 4.2: Detectable relative risk and disease allele frequency curves for 3-month KD response cohort, with 80% power, assuming a disease prevalence of 0.5%, an alpha of 0.007, 126 cases and control:case ratio of 1.4. Black line=recessive model; purple line=dominant model; orange line=co-dominant model

Taking the MAF of the variant with the lowest p-value in the 3-month KD response association analysis (MAF=0.017), approximately 800 individuals would be needed to achieve 80% power if the variant had a relative risk of 2.0, as shown in Figure 4.3. Over 3000 individuals would be needed if the variant had a relative risk of 1.5, and 4500 if the

variant had a relative risk of 1.4. This is assuming a co-dominant model; larger sample sizes would be required if a recessive model was assumed.

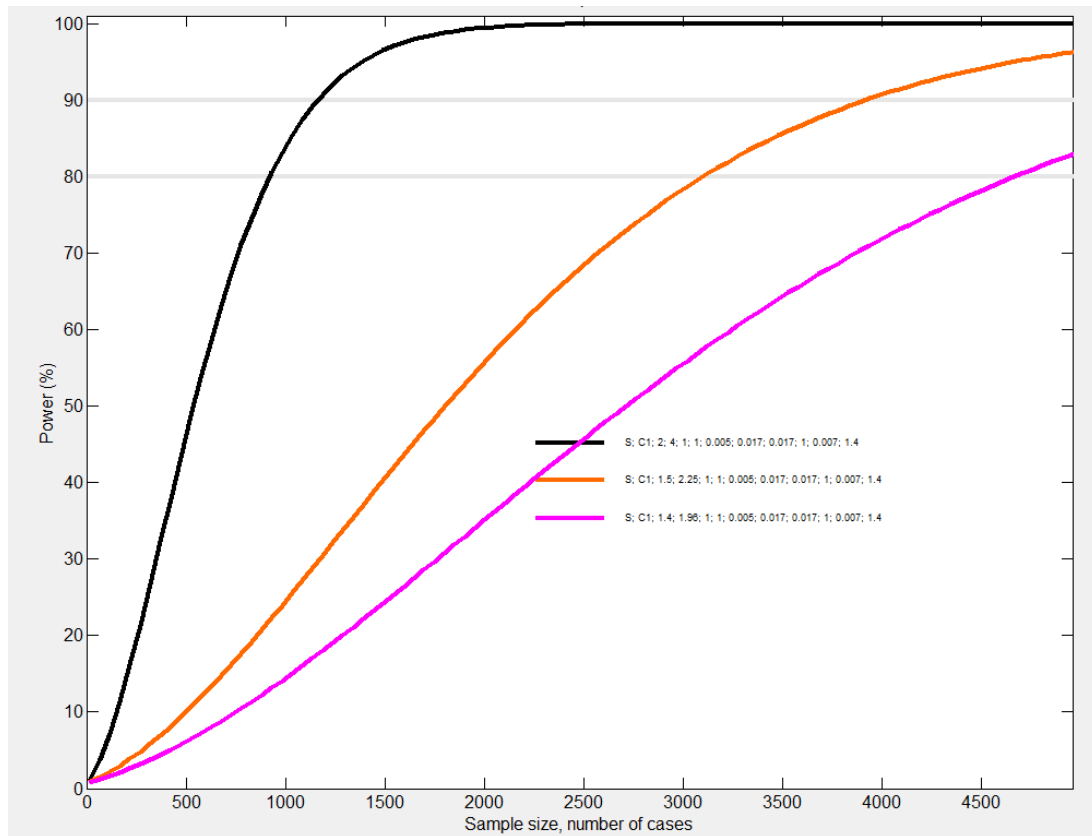


Figure 4.3: Power and sample size curves, using a co-dominant model, disease and marker allele frequencies of 0.017, assuming a disease prevalence of 0.5%, control:case ratio of 1.4 and alpha of 0.007. Black line=relative risk of 2.0; orange line=relative risk of 1.5; purple line=relative risk of 1.4.

4.4 Discussion

Key findings:

- The G allele of *KCNJ11* rs5216 (p.Leu267Leu) had the lowest p-value in the 3-month diet response association analysis, but it was not statistically significant when

adjusting for multiple testing (p-value obtained by permutation). This variant also had the lowest p-value when including self-reported ethnicity as a covariate, but the association was not statistically significant. The same variant did not have the lowest p-value in the summary KD response association analyses.

- No conclusions can be drawn from previously unreported variants in *KCNJ11* and *BAD* or those SNPs with MAF<0.01 in this cohort, due to the small sample size.
- Haplotype analyses (including all previously-reported variants in *KCNJ11* and *BAD*) yielded similar p-values to single SNP analyses. The haplotype with the lowest p-value in the 3-month response analysis contained rs5216 (p.Leu267Leu), rs5215 (p.Val337Ile) and rs5214 (p.Ser363Ser). The haplotypes with the 10 lowest p-values all contained rs5216 and rs5215.
- Replication is needed with a larger cohort.

The lack of statistically significant results in the candidate gene analysis is likely due to low power - the probability of successfully detecting a genotypic-phenotypic association. Power depends on the effect size of the variant, allele frequency, sample size, level of statistical significance set [584] and the type of assumed association (dominant, recessive or co-dominant), which is unknown. Due to the nature of dietary interventions such as the KD, the sample size in this study is necessarily small, at least in terms of genetic analyses.

Approximately 250 patients were following the KD for epilepsy in the UK in 2011-2012 and 264 started the KD in 2012-2013 (personal communication with Katherine Lord, Head of Nutrition and Dietetics, Southmead Hospital, and Chair of Ketogenic Diet Professional Advisory Group, UK). Further international collaboration would be needed to obtain a larger cohort size. If this were achieved, there would be more power to detect associations from variants with smaller MAF and lower effect sizes.

Appropriate measures were taken to account for multiple testing (although the multiple testing burden in these analyses was not great due to the small number of variants) and population structure. The variants with the lowest p-values in the association analyses still had the lowest p-values when including ethnicity as a covariate, which suggests that the results in the original association analyses were not confounded by population stratification. Standard methods of dealing with potential population stratification would have been preferable, but were problematic in this cohort: matching cases and controls in terms of ethnic group would have been practically difficult and self-reported ethnicity is not always accurate, as individuals may not be aware of their own ethnic admixture; inclusion of family-member controls was not possible as a family-based study design was not used; use of a panel of polymorphic markers that were unlinked to the candidate gene to test for stratification, was not possible as genome-wide SNP data were not available for all cases [585].

No SNPs were excluded from analyses based on predictions of functional consequence, due to the lack of certainty associated with classifying variants as functional or non-functional. In genetic association analyses, variants in coding regions tend to be prioritised, as they have been more consistently associated with Mendelian traits than have variants in intronic and regulatory regions, and are the least common types of polymorphisms [586]. However, synonymous or non-exonic variants can also affect gene function, although, when looked at singly, they are generally less likely to be associated with disease [586, 587]. The extent to which non-coding variants or variants not predicted to be of high functional importance (for example, those located in unconserved regions) contribute to complex diseases or traits is unknown. Variants predicted to be functional (or non-functional) by a single prediction algorithm are likely to have high false-positive rates [588], which is why scores from multiple algorithms were used in this study. Sole reliance on

functional prediction algorithms to include/exclude variants for analysis was not considered wise; for example, just because a variant is classified as a polymorphism, or it is present in dbSNP, that does not guarantee lack of association with a disease or phenotypic trait, as is the case with rs5219, rs5215 and rs5218. In this study, known information about variants is used to aid interpretation of results, for example, to determine whether top associated variants have a biologically plausible function with regards KD response.

4.4.1 Potential consequences of SNPs with lowest p-values

p.Leu267Leu merits further examination, due to the fact that it had the lowest p-value (even if not statistically significant) in the single-SNP association analysis and was present in the haplotypes most closely associated with 3-month diet response. The potential significance of p.Val337Ile and p.Ser363Ser (particularly when present together with p.Leu267Val), which also occurred in the haplotypes with the lowest p-values, is also of interest, as well as rs8175351 (p.Lys381Lys), which had the lowest p-value in the summary KD association analysis.

In pancreatic β cells, the binding of adenine nucleotides to Kir6.2 pore closes the channel; thus, variants affecting this subunit could influence the open-channel probability and alter cell excitability [589]. *KCNJ11* is expressed in many other human tissues aside from pancreatic β cells, including smooth and skeletal muscle, the brain, cerebellum, thyroid, adipocytes and the heart (www.genecards.org). If any of the variants of interest were to alter K_{ATP} open-channel probability in the brain, this may have implications for KD response. Considering the 'K_{ATP}-glycolysis hypothesis' [295], a mutant Kir6.2 subunit may affect the activation or opening of K_{ATP} channels in the presence of ketone bodies and/or reduction of ATP in the submembrane compartment, resulting in a failure to depress neuronal excitability.

The functional consequences of p.Leu267Leu and p.Lys381Lys on K_{ATP} channel activity are unknown. These variants are located in a conserved region and are rare in the general population, indicating that they influence gene/protein function and may be damaging: purifying selection acts against deleterious variants [555]. Synonymous variants may influence mRNA splicing (by creating new 'cryptic' splice sites or, indirectly, by affecting exonic splicing enhancers or silencers), leading to protein isoforms with different functions, mRNA stability or secondary structure (which affects splicing and processing of pre-mRNA), with downstream effects on protein expression and phenotypic traits [590-592]. Synonymous variants may also affect the tertiary structure of a protein or its function, which are not solely determined by the amino acid sequence of a protein, but also by translation rate; rare codons may cause pauses in ribosomal translation and common codons may speed up translation [593], and so changes in codon usage may influence protein folding and cause change in protein conformation and activity [591, 592]. No functional studies were found examining the effects of p.Leu267Leu or p.Lys381Lys.

Some studies indicate that a particular haplotype may have more functional significance than a single SNP. For example, the presence of both homozygous p.Glu23Lys, p.Val337Ile and heterozygous p.Leu270Val has been associated with higher insulin sensitivity index, compared with the combined wild-type genotypes in humans [594]. p.Leu270Val (alone, or in combination with other polymorphisms) had no effect on insulin secretion after intravenous glucose load or on 'glucose effectiveness' (undefined in the publication), although individuals who were heterozygous for p.Leu270Val and compound homozygous for p.Val337Ile and p.Glu23Lys had on average 62% higher insulin sensitivity index, compared to wild type carriers [594]. In theory, haplotype analysis could help determine whether the suggestive association between p.Leu267Leu and KD response is truly driven by this SNP alone and/or localise a region that contains other influential variants. Indeed,

results suggest that p.Leu267Leu in combination with p.Val337Ile and p.Ser363Ser is more closely correlated with 3-month KD response (due to the slightly lower p-value) than p.Leu267Leu alone.

The effects of p.Ser363Ser, a rare synonymous variant located in a non-conserved region, are unknown. The missense variant p.Val337Ile is not predicted to be of functional consequence according to SIFT, PolyPhen2, LRT and MutationTaster classification. This variant alone has not been associated with Type II diabetes [595] and it does not influence the properties of reconstituted human SUR1/KIR6.2 channels (including expression rate, single channel conductance, spontaneous open probability, and nucleotide and drug sensitivities [596, 597]. However, p.Val337Ile was tightly linked with rs5219 (p.Glu23Lys) in this cohort, as has been reported in other cohorts [594, 598, 599]. The high concordance between these two variants suggests that they may have originated in a common ancestor [595]. p.Glu23Lys has been associated with Type II diabetes in Caucasian [522, 599, 600], Indian [601] and Chinese/East Asian [602-605] subjects, although results are conflicting [606-608]. A meta-analysis of case/control studies investigating the relationship between *KCNJ11* polymorphisms and Type II diabetes found a significant result for p.Glu23Lys (odds ratio of 1.13); when stratifying results by ethnicity, a significant association was found in East Asians and Caucasians, but not among South Asians and other populations [609]. p.Glu23Lys is thought to have a diabetogenic effect by enhancing the open probability and reducing the sensitivity to ATP of the K_{ATP} channel, leading to the inhibition of insulin secretion [596, 610]. p.Glu23Lys has been shown to impair glucose-induced insulin release in humans [611] and individuals with EE (encoding two copies of glutamate) were found to have a higher disposition index (a marker of insulin sensitivity) than those with EK (one copy of lysine and one of glutamate) or KK (two copies of lysine) [612]. Others have not found differences in glucose-stimulated insulin secretion between carriers and non-carriers

of p.Glu23Lys [557, 613]. In one study, p.Glu23Lys was not associated with β -cell dysfunction or insulin resistance, but it was associated with diminished suppression of glucagon secretion in response to hyperglycaemia [614].

In addition to the putative effects of the variants of interest on ATP sensitivity or open probability of K_{ATP} channels in the brain, it is feasible that greater insulin sensitivity, leading to increased insulin-mediated glucose disposal (for example, into muscle [via GLUT4] or adipose tissue) may adversely affect KD response. For example, increased uptake of glucose into the liver (indirectly affected by insulin) would inhibit ketone body production [615]. It would be of interest to see if fasting plasma glucose, indices of insulin sensitivity (such as glycated hemoglobin - a measure of average plasma glucose concentration over prolonged periods of time) or β -cell function differ between carriers and non-carriers of certain *KCNJ11* variants in this cohort, or between all KD responders and non-responders.

If the same variants had generated the lowest p-values (or had the greatest difference between MAF in responders and non-responders) in association analyses conducted with 3-month and summary KD response, more credence would have been assigned to its putative effect on KD response. This was not the case, although the proportion of people classified as responders or non-responders in the 3-month and/or summary response categories was different; for example, somebody who was classified as a responder at the 3-month point may have subsequently had an increase in seizures and so would not be included in the summary KD group. In the 3-month analysis, any variant harboured by this individual would be associated with a favourable response but would not be associated with any sort of response in the summary analysis. Alternatively, it may be that certain variants only play a role in influencing KD response in the early stages of treatment, as blood glucose levels reduce and fatty acids are released by adipose tissue in order to allow effective conversion of fatty acids into ketone bodies by the liver. Gene-environment

interactions may also play a part; individuals who are overweight or obese may be more susceptible to impairment of glucose homeostasis due to *KCNJ11* variation (p.Leu27Leu or other variants) [595]. Variation in carbohydrate consumption may also affect to what extent such variation influences seizure control.

4.4.2 Rare and novel variation in *KCNJ11* and *BAD*

No conclusions can be drawn from the previously unreported variants or those SNPs with MAF<0.01 in this cohort. Based on its location in a highly conserved amino acid and predicted disease-causing or deleterious effect on gene function, NM_00525.3:c.[451G>A];[=], p.Val151Met is most likely to be of consequence. It was present in one non-responder, but more individuals with this variant are required to see whether it has a consistent effect on the phenotype.

Deep resequencing (Sanger sequencing of PCR-amplified DNA, where each nucleotide is read an increasing number of times during the sequencing process and reads are assembled against a reference genome) may be needed to detect rare, recent variants, not previously reported, for example, in dbSNP [616]. When the region of interest is limited (as in a candidate gene study), Sanger sequencing remains the gold standard, although deep resequencing with application of stringent genotype call probabilities may be the only way to identify the full extent of rare genetic variation in an individual [616].

4.4.3 Conclusions and future work

In conclusion, the candidate gene association study suggests that p.Leu267Leu in *KCNJ11*, possibly in combination with p.Val337Ile and/or p.Ser363Ser, may influence KD response at the 3-month point, and p.Leu267Leu with overall KD response, although results were not statistically significant when adjusting for multiple testing. It is possible that KD response is affected by K_{ATP} channel open probability and sensitivity to ATP, which influences insulin sensitivity (this is particularly plausible due to the linkage between p.Val337Ile and

p.Glu23Lys), but it unknown whether the variants of interest, either alone or in combination, have this effect on *KCNJ11* gene expression or protein function, in the brain or elsewhere. Even if they were to have this consequence, the effect sizes of the variants/haplotypes were not shown to be large, at least in this cohort. If a variant/haplotype was found, for example, in no non-responders and a high proportion of responders (no matter what time point response was measured at, or whether summary KD response was used), this would be of greater clinical value. Replication is needed in a larger, ideally ethnically-homogenous cohort, or at least one in which genome-wide SNP data were available for all participants.

The possibility cannot be discarded that other variants in *KCNJ11* or *BAD*, not found in this cohort, may play a role in KD response. Rare variants, which are predicted to have a larger phenotypic effect than common variants, are of particular interest.

GWAS results may aid selection of further candidate genes and potentially allow the candidate gene approach to be used on a genome-wide scale, as more SNPs (some located within or near candidate genes, or associated with them through LD) are identified throughout the genome [617].

5 Common genetic variation and response to the Ketogenic diet: a Genome-wide Association Study

5.1 Introduction

5.1.1 The genome-wide association study approach

For cohorts of unrelated individuals, GWAS are a widely-used method for identifying the genetic contribution to disease and treatment response. In contrast to candidate gene analyses, they may be conducted with no *a priori* assumptions, meaning that biological pathways not previously associated with a phenotype may be identified [500].

GWAS are based on the *common disease/common variant* hypothesis [618, 619], which states that common diseases are often due to combinations of alleles that occur frequently within the population [620]. Common diseases or traits tend to have multiple susceptibility alleles, each with a low effect size compared to those for rare disorders.

SNPs are used as genetic markers, exploiting the concept of LD - ‘the degree to which an allele of one SNP is inherited or correlated with an allele of another SNP within a population’ [464]. It is thought that a panel of 500,000-1,000,000 SNPs can capture >80% of common SNPs in European populations [621]. Phenotypic associations may thus be *direct*, where a genotyped SNP is shown to be statistically associated with the phenotype and this association is thought to be biologically plausible, or *indirect*, where a ‘tag’ SNP, which is in high LD with the potential SNP of interest, is shown to be statistically associated with the phenotype [464]. This concept is illustrated in Figure 5.1. The notion of indirect association may represent an advantage or disadvantage compared to candidate gene analyses; indirect association is an efficient way of obtaining genome-wide coverage with a subset of SNPs [622], but if the selected candidates are the primary functional genes with regards to

the trait in question, a GWAS may only be advantageous if the influential variant is actually genotyped [512]. Follow-up studies may be performed to investigate all variants in complete or near-complete LD with the tag SNP to identify causal variants [623].

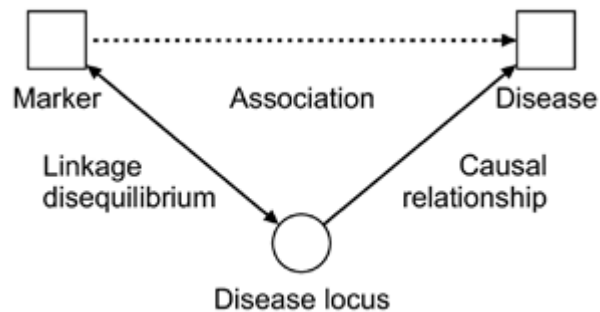


Figure 5.1: Illustration of the concept of direct or indirect association in GWAS

5.1.2 Quality control

In order to avoid bias (potentially leading to Type I or Type II error) arising from poor-quality SNPs and samples, a range of quality control criteria is customarily applied prior to GWAS analyses, on a per-SNP and a per-individual basis [510].

Commonly-used per-SNP quality control criteria include the following:

- SNPs with a certain proportion of missing genotypes, including those with significantly different missing genotype rates between cases and controls, are excluded in order to ensure accurate genotyping calls and to eliminate those with poor genotype probe performance [624, 625]. Different rates of missingness in case/control groups can lead to false associations.
- Deviation from HWE is principally used to identify gross genotyping error [626]. It is important to check deviation from HWE, as differences in allele frequencies in the two populations may be due to selection acting in different ways at a particular locus, rather than as a result of reproductive isolation of the populations. A fair comparison could not be

made between two populations if there were other factors involved, such as natural selection or non-random mating, that caused deviation from HWE in one population.

- When heterozygosity rates are higher than expected under HWE, this may indicate sample contamination, potentially due to interfering SNPs during hybridisation [573, 624, 627].

- Due to the need for very large sample sizes to detect associations from variants with low MAF, a threshold is usually set (such as 1, 2 or 5%, depending on the cohort size).

Furthermore, there tends to be more error with lower MAF SNPs in cluster-based calling algorithms, as it is more difficult to define a cluster with fewer observations [624, 625].

Commonly-used per-individual quality control criteria include the following:

- Exclusion of individuals with high or low proportions of heterozygous SNPs. Excess heterozygosity may indicate DNA contamination and low heterozygosity may indicate hybridization problems or inbreeding for that individual [573, 628].

- Gender mismatch between the phenotypic information collected and that imputed from genotyping data.

- Removal of duplicate and related persons; if the genotypes within a family are overrepresented, the cohort may no longer accurately reflect allele frequencies in the general population [510].

- Removal of persons with high proportions of missing data, which imply hybridisation problems, possibly caused by faulty arrays or poor DNA quality [573, 624, 625].

Quality control measures should not differ between different processing batches; if some aspect of the experiment is not randomised with respect to the phenotypic trait, this may lead to spurious associations.

Even after filtration of sub-optimal SNPs and individuals, manual inspection of cluster plots (particularly of top-associated SNPs) is still considered the best way to ensure genotype calls are robust [510].

5.1.3 Population stratification

There are various approaches for dealing with potential population stratification in GWAS. One is genomic control λ (λ_{GC}) [629, 630], defined as the ‘median χ^2 (1 degree of freedom) association statistic across SNPs divided by its theoretical median under the null distribution’ [631]. This approach corrects for population stratification, where necessary, by dividing the association-test statistic at each SNP by a common factor: the inflation factor [632].

Individuals may also be stratified according to genetic distance/relatedness with other participants. A common way of determining pairwise genetic distance/relatedness is by measuring identity-by-state (IBS). Pairwise differences in IBS (the genome-wide average proportion of shared alleles [558]) can be attributed to relatedness or to population-specific allele frequencies, according to geographic or ethnic groups [633]. IBS accounts for relationships up to $2N_c$ (where N_c is the effective population size) generations ago, as opposed to IBD, which depends on markers inherited from a recent common ancestor without recombination, and which only takes into account a limited number of generations within a known pedigree [634]. IBS can be used to demonstrate the pairwise genetic correlation between subjects and create a similarity matrix for principal component analysis (PCA) or the related multidimensional scaling (MDS). It is assumed that each population is defined by a characteristic set of allele frequencies, that each population is in HWE equilibrium, and that there is complete LD between loci within each population [635]. With PCA and MDS, each subject is placed in a point/vector in an X-dimensional space and the distance between each point reflects the relationships within the similarity matrix. The

genetic background of each subject is then represented by a vector of coordinates that reflects the genetic correlation relationships between subjects as closely as possible [631, 632]. The first vector represents the axis that accounts for the largest genetic variation; the second vector represents that which accounts for the second largest genetic variation, and so on [632]. A common way of accounting for population stratification in GWAS is to use these vectors as covariates, a technique employed by The Wellcome Trust Case Control Consortium [573]. Scree plots, where each vector is plotted against the proportion of phenotypic variation that it explains, are commonly used to decide how many vectors to include as covariates [636-638], as well as the plateauing of λ_{GC} [639]. Regions of long-range LD should be removed prior to computation of these vectors, as this could give an axis of variation which corresponds to genetic variation at that particular locus, rather than to genome-wide ancestry [631].

5.1.4 The multiple testing burden

Due to the fact that many markers are used in GWAS, the number of associations that could arise due to Type I error is large. The classic Bonferroni-corrected threshold of 5×10^{-8} or 1×10^{-8} , for an α of 0.05 and 0.01 respectively is often cited [640]. However, this wrongly assumes all markers are independent, and is thus overly conservative. Permutation is now considered the gold standard when adjusting for multiple testing in genetic association studies [562]. Phenotypes are randomly shuffled with respect to genotypes among observed data, removing any genotypic-phenotypic association; each random shuffle of the data represents one possible sampling of individuals under the null hypothesis of no association, a process which is repeated a certain number of times [464, 563].

A quicker and less computationally-intensive method is the calculation of the effective number of tests and then adjustment of the alpha threshold based on this number. There are various methods that have been proposed to calculate the effective number of tests,

such as the Cheverud-Nyholt method [641, 642] and K_{eff} (by Moskvina and Schmidt) [643]. The Cheverud-Nyholt method is based on the premise that, given a pairwise correlation LD matrix with eigenvalues, there are two extreme scenarios: all SNPs are perfectly correlated, or all SNPs are uncorrelated. The variance of the eigenvalues is used as a correlation measure for the effective number of SNPs. Moskvina and Schmidt's method tests, at every genetic locus, the null hypothesis that the distribution of the risk allele in the entire target population is equal to the distribution of the risk allele in cases, dependent on pairwise marker correlation coefficients, as well as on the significance level chosen for the individual tests.

Finally, false discovery rate has been proposed as a way to measure and control Type I error [644, 645]. Each marker is assigned a corresponding q-value instead of a p-value, which is a measure of significance in terms of the false discovery rate rather than the false positive rate (represented by p-values), that account for multiple testing. In this way, the researcher can control the expected proportion of false positives: for example, a threshold of 0.05 would mean that 5% of the observed results may be false positives [646]. However, q-values are determined from the overall distribution of p-values observed in the data and they are based on the assumption that the p-values are uniformly distributed under the null hypothesis of no association, which may not always be the case [643]. Moreover, false discovery rate is considered conservative for GWAS [647] and the original procedure proposed by Benjamini and Hochberg [645] is only valid when the tests are independent [648].

5.1.5 Interpreting results

One major difficulty with GWAS lies in the interpretation of results. Only 7% of SNPs previously found to be associated with a disease or trait in GWAS are located in protein-

coding regions [649, 650], and significant associations do not guarantee a causal effect. In this way, GWAS are more observational than experimental studies.

The first GWAS success identified a polymorphism in the Complement Factor H gene to be a major risk factor for age-related macular degeneration [651]. Since then, many associations have been reported, of which those within the context of pharmacotherapy and nutrition are of particular interest; for example, rs8014194, located within the *CLMN* gene, has been associated with variability in response to statins in terms of reduced total cholesterol [652]. Such genetic results have not commonly been assimilated into clinical practice, possibly due to small cohort sizes and/or the small effect size of identified variants, but they have certainly furthered understanding of the biology behind differential response to certain treatments and may help generate new hypotheses for further research.

5.1.6 Alternative approaches to the single-SNP method

The single-SNP analysis is a suitable approach when conducting the first GWAS for a particular phenotypic trait but, if one considers that a SNP is only a very small part of a larger biological web, the limitations of the single-SNP analysis approach when analysing complex diseases becomes clear. The effect of a variant on the phenotype may depend on the presence of other variants [653] or interactions with other interdependent components of the biological pathways within which the variant/gene plays a role [654]; it may be a combination of many factors that contributes to the phenotype. Alternative methods of processing SNP data may increase power to detect phenotypic-genotypic associations. There are three main approaches that are adopted when using preliminary GWAS findings for subsequent analysis:

- i) pooling information from multiple GWAS in a meta-analysis.

This increases the chances of separating true positives from false positives and is said to increase power [655-657].

- ii) modelling of epistatic interactions – association tests conditional on interactions, accounting for the effects that alleles of one SNP have on the effects of alleles of another SNP on the phenotypic trait.

This approach is based on the premise that stronger associations are revealed when genes interact. GWAS that account for SNP-SNP interactions may be more powerful than single-SNP GWAS in certain circumstances, and power is strongly affected by allele frequency of the susceptibility variants and by the extent of LD between tagged and typed variants [658]. However, interaction effects are even more difficult to detect than main effects and so if a GWAS is underpowered to detect associations from main effects, there is very little chance of detecting interaction effects if all possible pairwise interactions are considered [655]. It is common to consider only those pairwise interactions with a significant result, although this may mean that variants with modest effect sizes and large interactions may be missed [658].

- iii) pooled-variant approaches: most commonly pathway-based analyses, where single SNP GWAS results are grouped together in a known biological pathway to test whether the pathway is associated with the phenotypic trait [655], as well as gene-based analyses where SNPs are grouped into genes.

Pathway analyses are based on the premise that a SNP with too small an effect size to be associated with the phenotype by itself may have a greater effect when aggregated (in combination with other variants within the same unit) into a gene or pathway. This is particularly appropriate for complex traits, which may result from accumulated effects of genetic variants within genes or pathways. Variant-pooling approaches, particularly

pathway-based analyses, account for genetic heterogeneity: although affected individuals may share the same disrupted pathways, the mutated genes or variants within those pathways may differ amongst individuals. This approach also accounts for the fact that genes, particularly those present in a particular pathway, interact with other components [659]. Variant-pooling analyses may provide greater power to detect an association, if several of the SNPs within genes, or several of the genes within a pathway influence the phenotype; if only one SNP or gene has an effect, including others in the test will reduce power to detect an association [660].

Gene- or pathway-based analyses were first used to identify common themes amongst differentially expressed genes in gene expression datasets. There are several difficulties with using GWAS data for these analyses. Firstly, with gene expression data, each gene is represented by one value (the extent of over- or under-expression), but with GWAS, each gene is represented by a variable number of SNPs [661]. Secondly, standard genotyping platforms select SNPs based on LD patterns to provide maximum coverage of the genome, rather than coverage of all genes. Therefore, genes or pathways with more variants typed are more likely to be detected in association studies. Furthermore, aside from SNPs that are present in coding or regulatory regions, assigning SNPs to genes is difficult as SNPs may be in LD with multiple genes. There are several programmes that facilitate the transition of SNP lists to gene or pathway lists, which take into account LD (genes that are in high LD with a group of variants with low p-values, or with genes already known to be associated with the phenotypic trait, are more likely to be represented in the list of 'most associated' genes), such as ProxyGeneLD [662], INRICH [663] or Prioritizer [664]. However, there are no standard rules for assigning non-coding SNPs or those not in regulatory regions to genes and no single method has been shown to outperform others in analyses of different GWAS datasets [655]. There is less ambiguity when assigning variants to genes when using exome

chip genotyping arrays (or combination arrays, as used in this study) and so gene/pathway analysis may be a powerful approach. These are different to standard arrays, as they only include coverage of exonic regions; they aim to type putative functional variants – mostly non-synonymous variants but also other variants, such as those located in splice sites, promoter regions and extended major histocompatibility complex regions. In many cases, the functional interpretation will be easier, although this may not necessarily be the case for variants in regulatory regions.

Pathway-based analyses fall into two categories: those that test the competitive hypothesis, which assumes that genes with a particular function or involved in a particular pathway appear at the same frequency as in the reference genome, and those that test the self-contained hypothesis, which assumes that there is no association between the target gene set and the phenotypic trait [666].

One of the most popular methods that tests the competitive hypothesis is the GSEA algorithm [667], originally developed for gene expression data. Variants are assigned to genes (SNPs that are 500Kb away from any gene are discarded), and each gene is given a statistical value, which is equal to that of the variant assigned to the gene with the maximum X^2 value or the lowest p-value. Enrichment scores are calculated based on the extent to which the target gene set falls within certain pathways and normalised to account for gene size bias. The significance of the enrichment score for each pathway is estimated through permutation. Gene-wide p-values may also be obtained from programmes such as ProxyGeneLD [662], which assigns unadjusted p-values to each gene in the same way as GSEA (the SNP with the lowest p-value assigned to that gene becomes the gene-wide p-value) and then adjustment is made by multiplying that p-value by the number of SNPs assigned to the gene [668]. Another option is GSEA-SNP [669], in which variants are ranked according to results from single variant association tests and a group of variants, for

example within a certain pathway, is taken as a set. A variant-level enrichment score is calculated, which reflects the degree to which the variant set is over-represented at the top of the list of ranked variants. Whichever method is adopted, once a list of nominally-significant genes associated with the phenotype has been created (for example, those with $p\text{-value} < 0.05$), instead of considering genes one by one, one can determine whether a particular biological function is represented in this list more than what would be expected by chance. Programmes such as Database for Annotation, Visualization and Integrated Discovery (DAVID) [670], Protein ANalysis THrough Evolutionary Relationships (PANTHER) [671], and Ingenuity Pathways Analysis (IPA, from QIAGEN), facilitate enrichment analysis. After creating a list of nominally significant genes, this list can be uploaded into such programmes, some of which are available online, and enrichment scores calculated. Competitive analyses may not necessarily increase power in association studies, but serve to determine whether there is any biological function in common among all nominally-significant genes associated with the phenotype.

Another approach is to group variants into genes and conduct an association study testing the null hypothesis that no particular gene is associated with the phenotypic trait. An immediate generalisation of gene-based tests can then be conducted by grouping these genes into pathways. Examples of methods that test this self-contained hypothesis are the SNP Ratio Test [672] and the set-based test implemented in PLINK [558]. The SNP Ratio Test builds on the ratio of significant SNPs (as defined by a threshold, for example, 0.05) in a pathway and estimates the significance of the ratio via permutation. With the PLINK set-based test, the user must specify sets, which may genes or pathways, including which SNPs are included in which set. A group of independent SNPs within each set are identified, based on LD, and the statistic for each set is calculated as the mean of these single SNP statistics. Empirical p-values are then calculated, which correct for the number of

independent SNPs within a set. The advantage of the self-contained approach, compared to the competitive approach, is that there is no danger of underestimating the importance of a pathway because variants/genes that contribute to the effect of a particular phenotype on a pathway but that, individually, have less of an effect size than other variants/genes, are not included (as may occur if only the most significant SNPs were considered) [661].

Gene-or pathway-based analyses have identified significant genotypic-phenotypic associations, even when no individual SNP passed genome-wide significance thresholds [661]. A multiple-component approach has been recommended when analysing GWAS data [673].

A GWAS will be the first step in attempting to identify whether common genetic variation (in genes aside from *KCNJ11* and *BAD*) influences KD response. A single-SNP, gene-based and pathway-based approach will be adopted. This may help elucidate the mechanisms behind the antiepileptic effects of dietary treatment.

5.2 Methods

5.2.1 Phenotypic data

Two approaches were adopted:

- i) Using KD response at 3-month follow-up.

A binary phenotype was adopted, using seizure reduction thresholds that are considered clinically relevant to the population in question: $\geq 50\%$ seizure reduction=responder; $\leq 50\%$ seizure reduction=non-responder. 3 months after diet onset is the first time point at which KD response is formally assessed in the KD clinic and the greatest number of individuals in this study has KD response data available at this point.

- ii) Using a summary of KD response over time.

KD response is not always consistent over time, as shown in Chapter 3, and a summary of longitudinal response attempts to account for this variability, rather than taking response at one specific point in time. Individuals who achieved $\geq 50\%$ seizure reduction at every time point recorded were classified as responders, and those who achieved $< 50\%$ seizure reduction at every time point recorded were classified as non-responders. Thus, this category serves as a measure of long-term (consistent) response. Cases who consistently reported approximately 50% seizure reduction over time, or those with very fluctuating diet response over time were excluded from the summary diet response classification, as responder/non-responder status was ambiguous.

This GWAS is primarily concerned with common genetic variants, which are predicted to have a small effect size, and so extreme KD response (defined as $\geq 75\%$ seizure reduction) was not used as a phenotype.

5.2.2 Genetic data

DNA sample concentration was measured with NanoDrop ND-1000 Spectrophotometer (Thermo Fisher Scientific Inc., USA) and then diluted to 6 μ l at 50ng/ μ l. Samples were sent to AROS Applied Biotechnology A/S, Denmark to be genotyped with the *Infinium HumanOmniExpressExome Beadchip* (Illumina Inc, San Diego, USA) in two batches. See Figure 5.2 for numbers of individuals recruited and included in the final GWAS.

The Illumina chip has 730,525 common variants at $> 5\%$ MAF and $> 240,000$ functional exonic markers, selected from $> 12,000$ exome and genome sequences. Through LD ($r^2 > 0.8$), it provides coverage for 73% MAF $> 5\%$ and 58% MAF $> 1\%$ outside coding regions from the 1kGP European cohort (Human OmniExpress BeadChip Data Sheet, www.illumina.com) and 10% MAF $> 5\%$, 9.6% MAF $> 2.5\%$ and 8.8% MAF $> 1\%$ ($r^2 > 0.8$) within exonic regions (HumanExome BeadChips Data Sheet, www.illumina.com).

Raw genotyping data was analysed by Dr Costin Leu, using Illumina GenomeStudio software (v2011.1, Illumina Inc, San Diego, USA) with the Genotyping module (v1.0, Illumina Inc, San Diego, USA). PLINK (v1.07, <http://pngu.mgh.harvard.edu/purcell/plink>, [558]) was used for quality control and to conduct the GWAS.

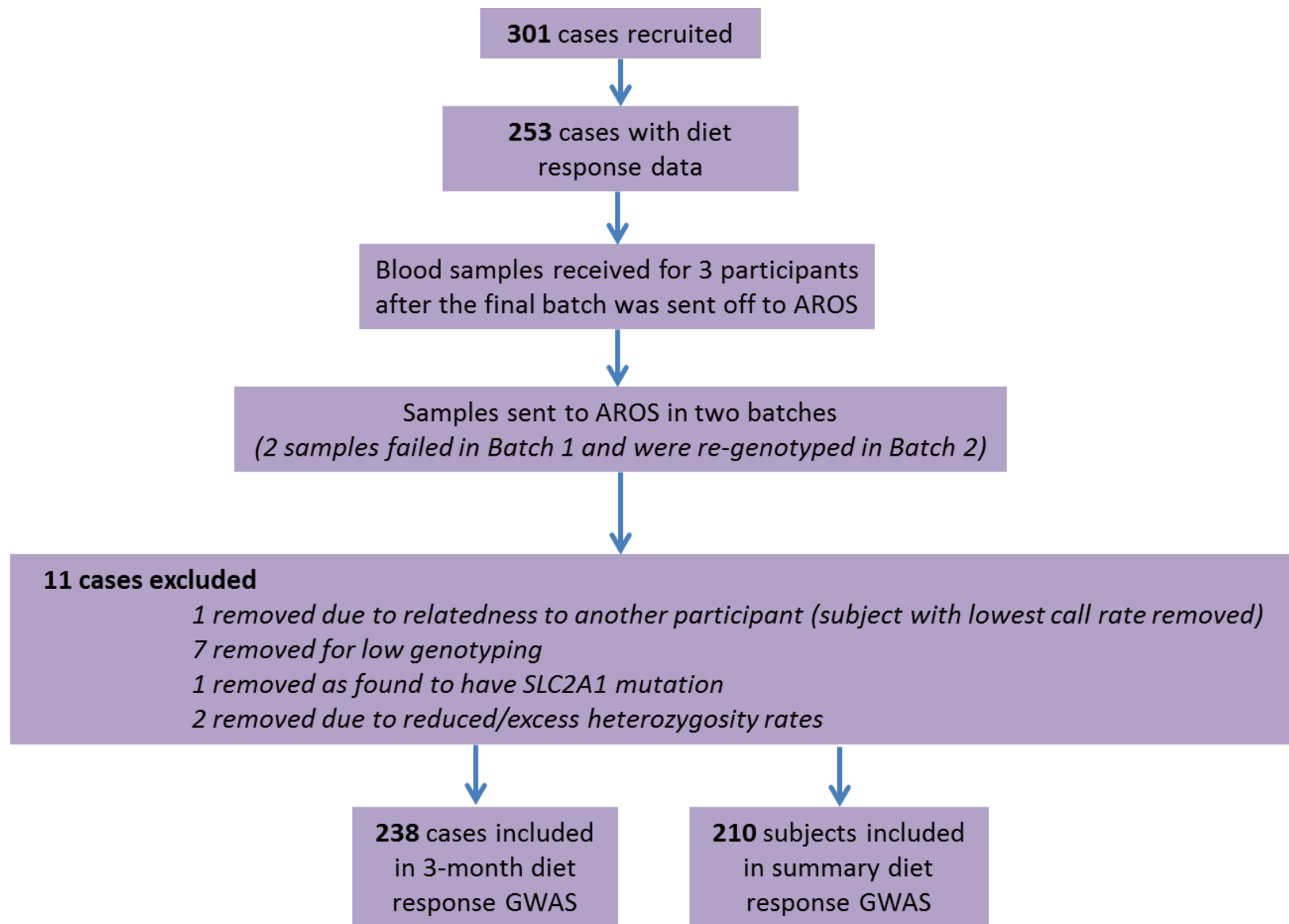


Figure 5.2: Flowchart of cases recruited and included in final GWAS

5.2.3 Quality control

All DNA samples were prepared to 6µl at 50ng/µl to ensure consistently high DNA quality/concentration.

All markers and individuals had to pass several tests to be included in the GWAS, with the aim of excluding poor-quality SNPs and samples without removing genuine phenotype-genotype associations. Guidelines on such quality control criteria have been outlined elsewhere [510] and implemented in previous epilepsy GWAS [582].

5.2.3.1 Per-individual quality control exclusion criteria

- i) Individuals with an excessive (>0.33) or reduced (<0.25) proportion of heterozygote genotypes.

The observed number of homozygous genotypes and total number of non-missing genotypes were calculated in PLINK (command: `--het`). The proportion of heterozygous SNPs for each individual, stratified by self-reported ethnicity, was plotted in Microsoft Excel 2010 (v. 14, Microsoft, Washington, USA), as shown in Figure 5.3. These ethnic groups reflect the range of ethnicities in this cohort, bearing in mind that allele frequency differences between different continents are much greater than those within continental subgroups [639] and that population admixture also affects allele frequency due to influx of alleles from other subpopulations [674]. Visual inspection of this plot enabled the researcher to eliminate outliers in each ethnic group (outliers are marked within circles in Figure 5.3).

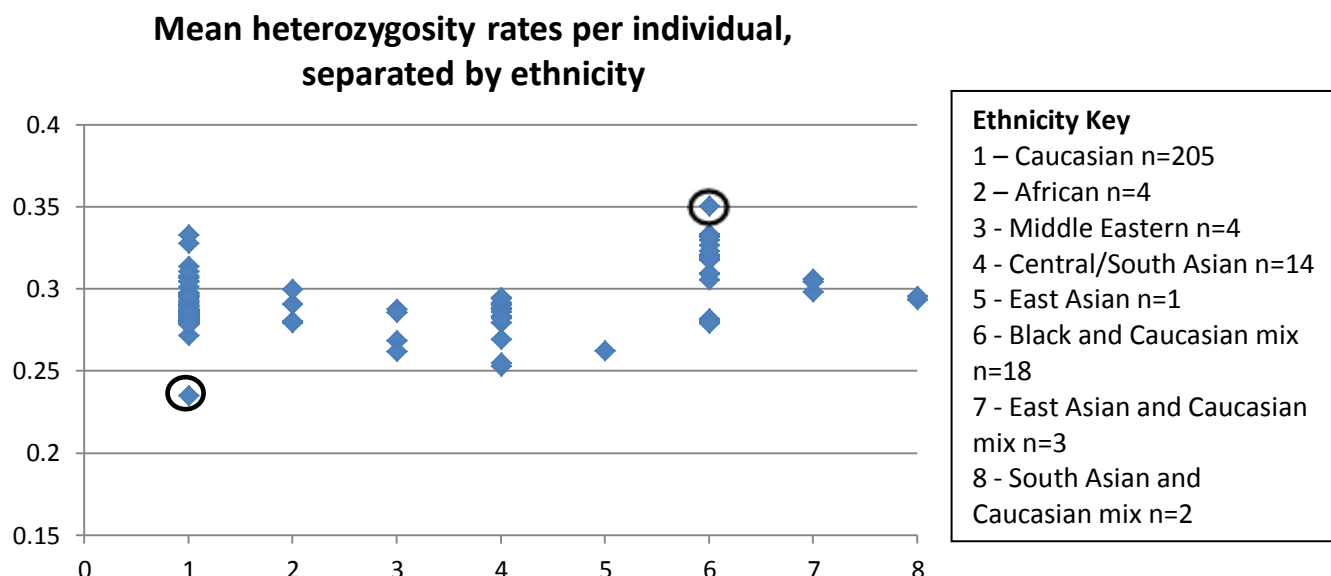


Figure 5.3: Mean heterozygosity rates per individual, separated by ethnicity

- ii) Mismatch between gender reported in the phenotype file and gender imputed from PLINK (command: `--check-sex`).

Homozygosity estimates were calculated across all X-chromosome SNPs for each individual; a male call was made when homozygosity estimate > 0.8 and a female call was made when homozygosity estimate < 0.2 .

- iii) Duplicate and related persons.

IBD was calculated between all pairs of individuals ($\hat{\pi}$ [proportion of IBD] > 0.4 ; command: `--genome`). Full siblings share an average of 0.498 genome-wide IBD, with a range of 0.374–0.617 [675]. A $\hat{\pi}$ threshold of 0.4 is therefore thought to remove duplicated cases and full sibling pairs. Where related individuals were found, the case with the lower call rate was removed.

- iv) Individuals with $> 2\%$ missing SNP data (command: `--mind 0.02`).

5.2.3.2 Per-SNP quality control exclusion criteria

- i) Genotyping rate <0.02 (command: `--geno 0.02`), following manual re-clustering of SNPs with $>1\%$ 'no calls' in GenomeStudio (clustering within Batch 1 did not appear to differ from Batch 2).

'No calls' refers to the number of genotypes in each sample with a GenCall score (a quality score indicative of the reliability of each genotype call) below 0.15 - genotypes with a GenCall score <0.15 are not assigned genotypes as they are considered to be too far from the centre of the cluster to which the data point is associated to make reliable genotype calls (TECHNICAL NOTE: ILLUMINA® DNA ANALYSIS. Infinium® Genotyping Data Analysis: A guide for analyzing Infinium genotyping data using the Illumina GenomeStudio® Genotyping Module, www.illumina.com).

- ii) Cluster separation values <0.3 and Het-excess values between 0.1 and -1 and between 0.1 and 1, given in GenomeStudio.

These exclusion criteria avoid SNP calling errors from the re-clustering steps. Higher cluster separation values (a measure of how close the raw colour signal data corresponds to each allele when assigning genotypes; a value close to one indicates homozygosity for one allele, a value close to 0 indicates homozygosity for the other allele and 0.5 indicates a heterozygous genotype) and Het-excess values (a measure of excess heterozygote calls relative to the expectation under HWE; Het-excess = -1 when no heterozygotes are called and =1 when all genotypes are called heterozygotes) closer to 0 indicate more reliable genotype calls.

- iii) MAF <0.05 (command: `--maf 0.05`).

Some previous GWAS have set the MAF threshold at 0.01 (including common SNPs and those of intermediate frequency) [582, 676, 677] but, as outlined in Section 5.3.5, this

study is already underpowered and so it would not be worth including variants with much smaller MAF.

- iv) Deviation from HWE, $p\text{-value} < 1 \times 10^{-6}$, obtained from an exact test [678]
(command: `--hwe-all --hwe 0.000001`), a threshold that has been used in many previous GWAS [625, 679-682].

This threshold was principally set to identify gross genotyping error, for which extreme deviation from HWE is typically used [626]; it should not be used for identifying minor genotyping error, as deviation from HWE has low sensitivity for testing for genotyping error [683-685]. The exact test method was used for testing deviation from HWE because it controls Type I error rates within large and small cohorts and is accurate for rare and common allele frequencies [678]. In contrast, deviation from HWE calculated using the chi-squared test may be inaccurate for small sample sizes or rare alleles; there may not be enough individuals in the sample to adequately represent all genotypes, and so the asymptotic assumption of the chi-squared distribution (the sum of k independent random variables with finite mean and variance, which converges to a normal distribution when k is of a sufficiently large size) no longer holds.

5.2.4 Covariates and population structure

As outlined in Chapter 3, the effect of a range of demographic and clinical factors on KD response was assessed by logistic regression.

IBS distance was calculated between all pairs of individuals in PLINK, using the commands `--genome`, and `--read-genome --cluster`. Known regions of high LD (see Table 5.2 in Appendix 5.1) were excluded and the dataset was thinned to create a set of independent SNPs, using the command `--indep-pairwise 1500 150 0.2`. With this command, all SNPs were pruned in a sliding window of size 1500 and were required to have $r^2 < 0.2$. It is recommended to

exclude long-range LD regions and perform LD pruning [686], as some regions of the genome may be over-represented due to high LD and these regions may bias clustering: vectors that capture population differentiation (such as those used in PCA and MDS) may only reflect genetic variation at those specific loci [631, 687].

Self-reported ethnicity or country of origin could have been used to account for ethnicity (for example, with a stratified Cochran-Mantel Haenzel test), but using IBS to determine genotypic differences between participants has the advantage that genotyping data may distinguish between people who reported themselves to be from a specifically-defined population, or reveal population admixture. Pairwise IBS distance was visualised by performing MDS analysis on the matrix (PLINK command *--mds-plot*) and plotting the value of the position on the first vector against the value of the position on the second vector for each individual in Microsoft Excel 2010 (v. 14, Microsoft, Washington, USA). This produced a scatter plot where each point represents one individual, giving a two-dimensional reduced representation of the data. As shown in Figure 5.4, self-reported ethnicity did not always correlate with genotypic similarity.

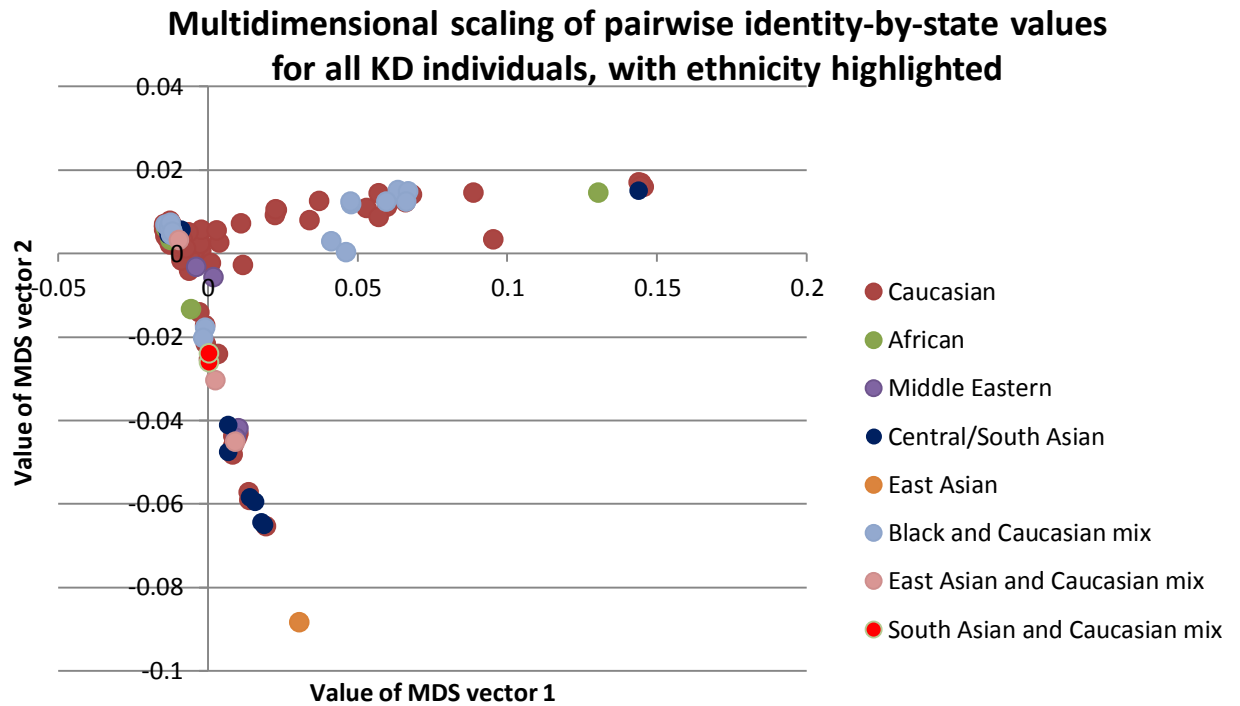


Figure 5.4: For all individuals with Ketogenic diet response data: plot of values of the first and second vectors obtained by multidimensional scaling of pairwise identity-by-state values, with ethnicity highlighted

The MDS vectors were also plotted with responders and non-responders highlighted in order to determine whether there were any systematic differences in the data – this would be shown by any clustering of responders/non-responders. These differences may not only be due population structure, but also other factors that may confound association analyses such as cryptic relatedness or batch effect. Response status at the 3-month point was taken, as most individuals had response data for this time point.

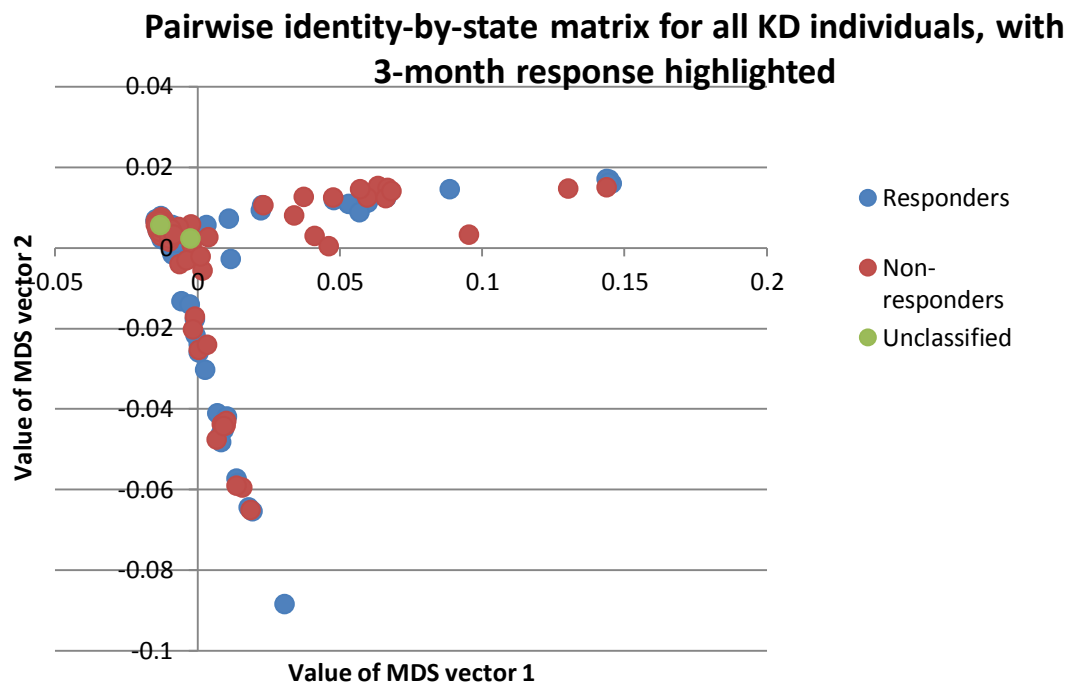


Figure 5.5: For all individuals with Ketogenic diet response data: plot of values of the first and second vectors obtained by multidimensional scaling of pairwise identity-by-state values, with 3-month response highlighted

As shown by Figure 5.5, responders and non-responders overlap and are virtually indistinguishable from each other, indicating that the phenotypes are appropriately randomised.

In order to determine the extent to which these vectors account for phenotypic variability in this cohort, a scree plot (see Figure 5.6) was plotted in R: A Language and Environment for Statistical Computing (v. 3.0.0, R Foundation for Statistical Computing, Vienna, Austria). Using the 'ibs()' and 'cmdscale()' functions, pairwise IBS values were calculated and MDS analysis performed on the distance matrix, just as in PLINK. Eigenvalues (scalar multipliers of the vectors) were then calculated for each vector and a scree plot completed using the 'plot(cps.full\$eig[1:10]/sum(cps.full\$eig))' function ('cps.full' is the name given to the distance matrix).

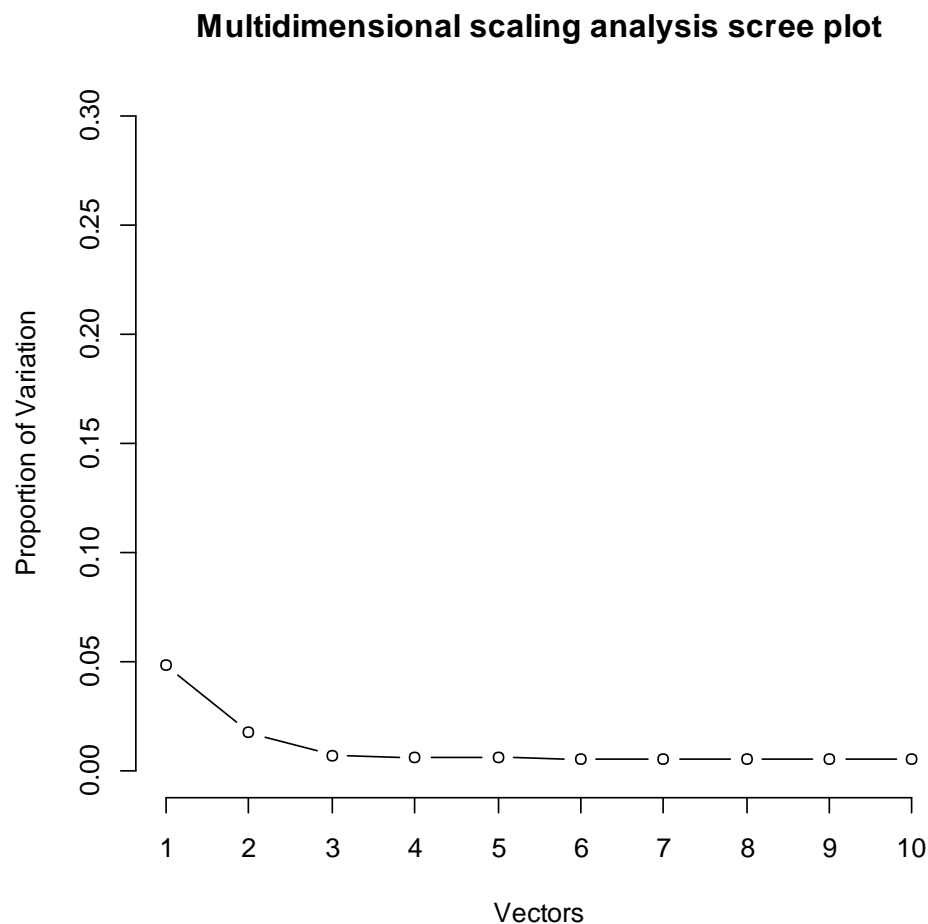


Figure 5.6: Scree plot for the first 10 eigenvalues of the Multidimensional scaling matrix

From this plot, one can see that the first vector only explains 5% of the total phenotypic variance; the second vector explains close to 0% of the variance, and the remaining vectors

explain even less. Based on these results, it may be considered reasonable to include the first component in association analyses but, as it accounts for so little phenotypic variability, the question is also raised of whether vectors need to be included in analyses at all. If more components are included in analyses than are necessary, noise is introduced into analyses [688]. As explained in the following section, analyses were completed with no adjustment for population stratification, and also including the first vector as a covariate. λ_{GC} [629, 630], was calculated in PLINK with the command `--adjust`, and used as a measure of bias due to population stratification in association analyses.

5.2.5 GWAS

An association analysis was performed using a Fisher's allelic association exact test (which compares frequencies of alleles in cases versus controls) in PLINK, with the commands `--fisher --adjust --ci 0.95`. This test is suited to small sample sizes and avoids bias due to distributional approximation.

A logistic regression additive model was also performed in PLINK (commands `--logistic --covar`), including the first MDS vector values as covariates.

To facilitate the creation of quantile-quantile plots, the command `--qq-plot` was added. Quantile-quantile plots are a commonly used tool to represent GWAS data and assess whether more significant results have been generated than expected by chance. If results deviate from the expected values across the whole distribution, this may be due to undetected population stratification, cryptic relatedness or batch effects [689]; variants with the lowest p-values that deviate from expected values correspond to those with the greatest effect on the phenotype.

5.2.6 Correction for multiple testing

PLINK was used to obtain p-values by permutation, with the commands `--mperm 1000`.

≥1,000 permutations are needed to estimate a 0.05 threshold [565, 566]. 1000

permutations were chosen in order to obtain precise p-values whilst remaining realistic with regards to time needed for these computationally-intensive permutation procedures. The significance threshold was set at 0.05, as is widely adopted [690].

An alternative (and less computationally-intensive) method of correcting for multiple testing was also used: the effective number of independent tests was calculated, according to Moskvina's method [643], by Dr Jon White. This method accounts for the pairwise LD of the typed SNPs and the significance level used in the individual tests. Moskvina's method is preferable to many other methods, as the tests remain independent even for markers that are moderately or highly correlated [643]. For example, both the Cheverud-Nyholt method and simple M overestimate the effective number of tests if large groups of markers are highly correlated [568, 691]. With simple M, if each chromosome is not broken down into smaller regions, analysis requires a great deal of computational memory and time; the accuracy of the threshold applied may depend on variations in region size [640].

Taking the significance threshold for a single test to be 0.05, the threshold adjusted for multiple testing (based on the effective number of tests) was calculated using the SIDAK method [692], by Dr Jon White: $(1 - (1 - \alpha)^{1/\text{number of tests}})$ [561].

A suggestive significance threshold was set by dividing 1 by the number of tests, a strategy used in previous publications [693, 694].

5.2.7 Visualisation of GWAS results

Haploview (v4.2, <http://www.broadinstitute.org/haploview>, [695]) was used to generate Manhattan plots and to visualise the LD pattern around the SNPs of most interest. Regional plots for the most significant SNPs were created using LocusZoom (<http://csg.sph.umich.edu/locuszoom>, [696]), with genome build/LD population hg19/1000 Genomes March 2012 EUR. LD was displayed as D'/LOD (D' , the normalised covariance for

a given marker pair, indicating to what degree two alleles have or have not been separated by recombination; LOD, logarithm of the odds, is a statistical estimate of whether two loci are likely to be located near to each other on a chromosome and are therefore likely to be inherited together). The LD Display Spacing in Haploview was adjusted so as to ensure the position of the SNPs lined up as closely as possible to the Locuszoom display. This does make the display of D' /LOD a little less clear, as the space between each coloured triangle is increased, but it facilitates visualisation of which gene(s) the SNPs are in LD with.

Quantile-quantile plots were created using R: A Language and Environment for Statistical Computing (v. 3.0.0, R Foundation for Statistical Computing, Vienna, Austria) and the commands *plot(-log(data\$QQ, 10), -log(data\$UNADJ, 10)* and *abline(a = 0, b = 1)*.

Genes located within the LD blocks of the top SNPs were looked-up in Genedistiller (v2, www.genedistiller.org, [697]) to obtain information regarding protein families (ENSEMBL), pathways (KEGG), Mouse Genome Database phenotypes and GeneOntology terms, with data based on NCBI genome build #37. Human phenotypic-genotypic associations were investigated using Online Mendelian Inheritance in Man (<http://omim.org>), The Database of Genotypes and Phenotypes (<http://www.ncbi.nlm.nih.gov/gap>), Genetic Association Database (<http://geneticassociationdb.nih.gov>, [698]) and Phenotype-Genotype Integrator (<http://www.ncbi.nlm.nih.gov/gap/phegeni>, [699]). These sites were used to determine whether the genes had been associated with a neurological or metabolic phenotype.

The top associated SNP was looked up in UCSC Genome Browser (<http://genome.ucsc.edu>) and HapMap Genome Browser release #28 (<http://hapmap.ncbi.nlm.nih.gov>) to search for physically close genes (including transcripts), SNPs and gene regulatory elements.

5.2.8 Power calculations

All power calculations were performed using PGA Power Calculator [581], taking the sample size/case:control ratio from the 3-month diet response phenotype model, as this was the largest cohort.

Curves depicting potential detectable relative risk and MAF in this cohort, with varying power, were plotted, as were power and sample size curves. An r^2 value of 0.9 between a genotyped marker and causal variant was assumed, and an alpha level according to the effective number of tests was used.

As there is no naturally-existing KD, or indeed calorie restriction, at least in Europeans, disease prevalence (or prevalence of the trait) is hard to gauge. The disease prevalence of epilepsy (0.5%) was therefore used, as in other studies [582].

The inheritance model is unknown and so power calculations were performed using co-dominant, dominant and recessive penetrance models.

5.2.9 Gene- and pathway-based analyses

Each variant present in the association analyses was assigned to a gene in wANNOVAR [574]. Variants in exonic regions, those that overlapped 5' UTRs (untranslated regions), 3'UTRs or introns, those within 2bp of splicing junctions, or within 1kb of a transcript start site were assigned to the corresponding gene (as per default setting for wANNOVAR annotation). Intergenic variants were discarded, as it is often difficult to determine with which gene in their LD region (if any) they interact and thus their functional consequence is hard to interpret.

A set file was created for use in PLINK, with a list of genes and the SNPs included in those genes. The PLINK set-based test [558], which has been found to be more powerful (although also slower) than other gene-based algorithms [700], was used to test whether

any particular gene(s) are associated with KD response (command `--set-test --set Genes.txt --mperm 1000` [where *Genes.txt* is the list of genes and SNPs included in those genes]). This self-contained approach, rather than a competitive approach was adopted in order to account for possible interactions from variants with less of an effect size. Furthermore, due to the low power of the study, a limited number of genes would be classed as nominally significant. As the gene list is determined by single variant tests, if these are underpowered, the power of enrichment tests will also be low as the number of nominally significant genes is low – the higher the number of genes in the list, the more likely they are to generate a low p-value [701].

The PLINK set-based test first determines a set of independent SNPs within each set, based on their LD relationship. It then conducts a single SNP analysis (in this case, a Fisher's exact test with the same quality control filtering as adopted as in the single variant GWAS) and selects independent SNPs within each set with a p-value below a pre-defined threshold. The statistic for each set is calculated as the mean of the p-values of each of these SNPs. Empirical p-values are created by permutation, which accounts for the number of multiple SNPs within a set but not for the number of sets

(<http://pngu.mgh.harvard.edu/~purcell/plink>). The reason for creating a set of independent SNPs is that p-values of SNPs within the same gene are often not independent, as SNPs may be in LD. P-values for a gene or pathway will therefore be inflated if there are large correlations among SNPs in the gene. The reason for setting a maximum number of SNPs within each set is that genes or pathways with more variants typed are more likely to show association by chance alone [702]. In this study, SNP independence was defined as $r^2 < 0.5$ (as per default settings), the maximum number of independent SNPs within a set was defined as 5 (as per default settings) and all SNPs with a p-value of < 1 were included.

A PLINK set-based test was also used to conduct a self-contained pathway-based analysis, as opposed to taking a proportion of genes from the gene-based test and conducting a competitive analysis, in order to account for possible interactions from variants/genes with less of an effect size. A list of genes and their associated pathways was downloaded from the Broad Institute website (www.broadinstitute.org/gsea). This list included KEGG, PANTHER, REACTOME, BIOCARTA and PID (Pathway Interaction Database) pathways. This avoids potential bias from limiting oneself to one set of pathways. Analyses were not conducted with GO (Gene Ontology) terms due to the general nature of many of these categories, which makes interpretation difficult. Another set file was created for use in PLINK with a list of pathways and a list of SNPs included in these pathways.

A Bonferroni-adjusted significance threshold was set by dividing 0.05 by the number of genes or pathways included in the analysis.

5.3 Results

5.3.1 Quality control

A breakdown of average SNP call rates by batch and plate is given below:

Batch 1, Plate 1 (11/06/12): average call rate=0.997. No samples had a call rate<0.990

Batch 1, Plate 2 (11/06/12): average call rate=0.995. Two samples had a call rate<0.990
(these were the two samples that failed)

Batch 1, Plate 3 (15/08/12): average call rate=0.993. Eight samples had a call rate<0.990

Batch 1, Plate 4 (15/08/12): average call rate=0.996. No samples had a call rate<0.990

Average call rate within Batch 1=0.995

Batch 2, Plate 1 (29/01/13): average call rate=0.993. Four samples had a call rate<0.990

These average call rates (all >0.990) are consistently high across all plates/batches.

Genotype concordance rate was not checked, as no samples were genotyped twice (with the exception of two samples that failed in Batch 1).

One subject was removed because of imputed relatedness to another study participant (a pair of siblings were recruited and the subject with the lowest call rate was removed).

Seven subjects were removed due to low genotyping rates. Two subjects were removed due to excess/reduced heterozygosity rates. One subject was removed due to presence of *SLC2A1* mutation (identified from whole exome sequencing, after recruitment and genotyping).

5.3.2 Covariates and population structure

There was no strong evidence to suggest that KD response was affected by any clinical or demographic factors (see Chapter 3) and so none were included as covariates in the GWAS.

λ_{GC} was 1.00 in the Fisher's exact test, both with the 3-month response phenotype, and with the summary response phenotype. In the logistic regression including the first MDS vector values as covariates, λ_{GC} was 1.008 in the 3-month KD response GWAS and λ_{GC} was 1.0002 in the summary KD response GWAS. This indicates absence of importance of population stratification, and that noise is introduced when including MDS vectors as covariates.

Quantile-quantile plots were also examined to determine whether results deviated from the expected values across the whole distribution (this may be due to undetected population stratification or cryptic relatedness) to a lesser extent in the Fisher's exact test, compared to the logistic regression.

Quantile-quantile plots of the genome-wide distribution of results for all GWAS (see Figure 5.7 and Figure 5.8) indicate deviation from the null hypothesis of no association only in the

upper tails, corresponding to the SNPs with strongest evidence for association. Consistent with the previous findings, slightly less bias is seen when using the Fisher's exact test – the line of black dots (each representing a p-value) is closer to the line representing the expected p-values.

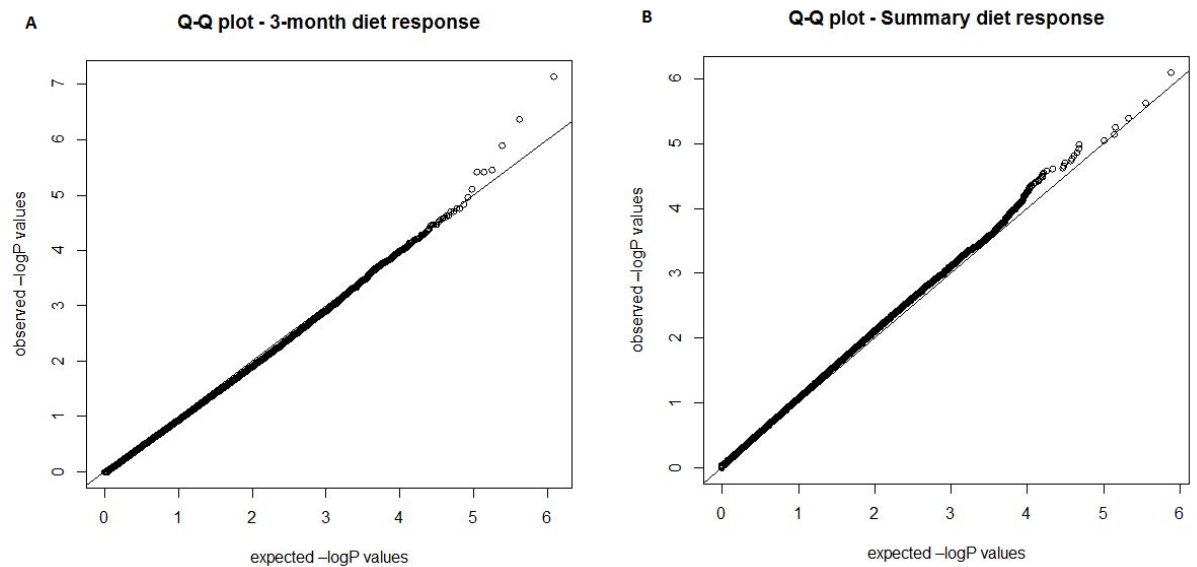


Figure 5.7: Quantile-quantile plots of GWAS results from Fisher's exact test, using unadjusted p-values. A) 3-month diet response, $\lambda=1.00$; B) Summary diet response, $\lambda=1.00$

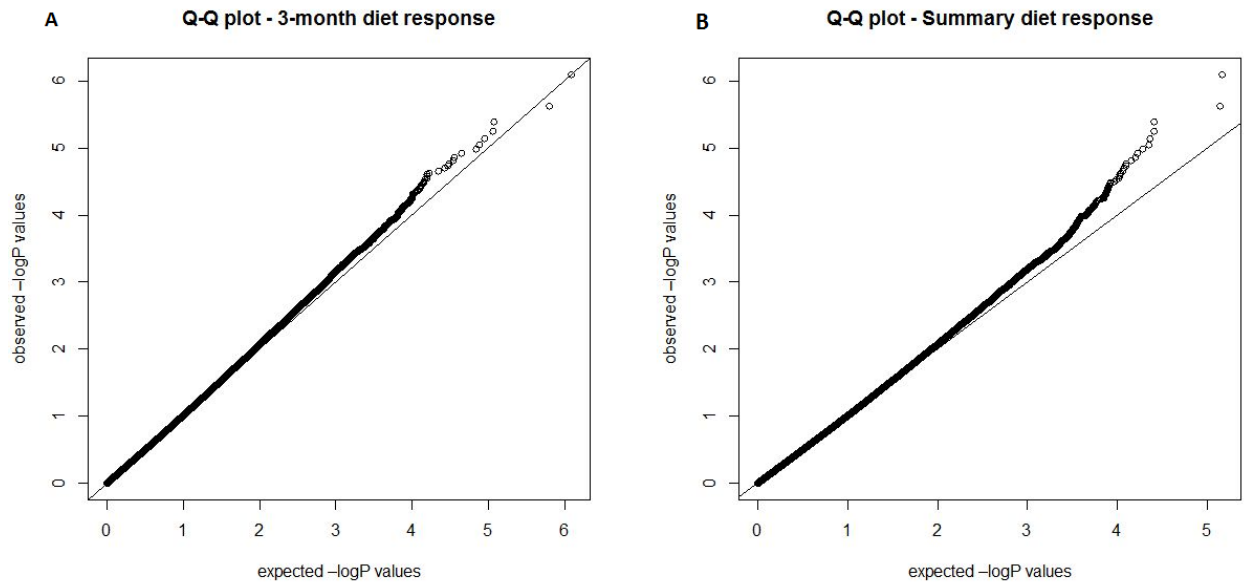


Figure 5.8: Quantile-quantile plots of GWAS results from logistic regression, using unadjusted p-values. A) 3-month diet response, $\lambda=1.008$; B) Summary diet response, $\lambda=1.0002$

Based on the above findings, the Fisher's exact test was used for association tests.

5.3.3 Correction for multiple testing

The threshold for genome-wide significant association was set at 0.05 for p-values obtained from 1000 permutations.

The alternative significance threshold was 1.01×10^{-7} ($\alpha=0.05$) and the suggestive significance threshold was 2.12×10^{-6} ($\alpha=1.00$), after SIDAK correction for testing 471,955 effectively independent SNPs.

5.3.4 GWAS results

Following quality-control filtering and exclusion of ethnic outliers, 614902 SNPs remained in each GWAS.

116 non-responders and 122 responders were included in the 3-month diet response GWAS; 100 non-responders and 110 responders were included in the summary diet response GWAS.

rs12204701 reached significance in the 3-month KD response GWAS (unadjusted p-value= 7.34×10^{-8} ; p-value obtained from 1000 permutations=0.039): the minor allele, A, was more frequent in non-responders than responders. rs12402021 (unadjusted p-value= 4.28×10^{-7} ; p-value obtained from 1000 permutations=0.17) and rs4696766 (unadjusted p-value= 1.29×10^{-6} ; p-value obtained from 1000 permutations=0.40) reached suggestive significance.

No SNPs reached significance in the summary KD response GWAS. rs4674639 reached suggestive significance (unadjusted p-value= 1.31×10^{-6} ; p-value obtained from 1000 permutations=0.40).

Regional association plot and LD maps for intergenic SNPs of significance or suggestive significance are given below.

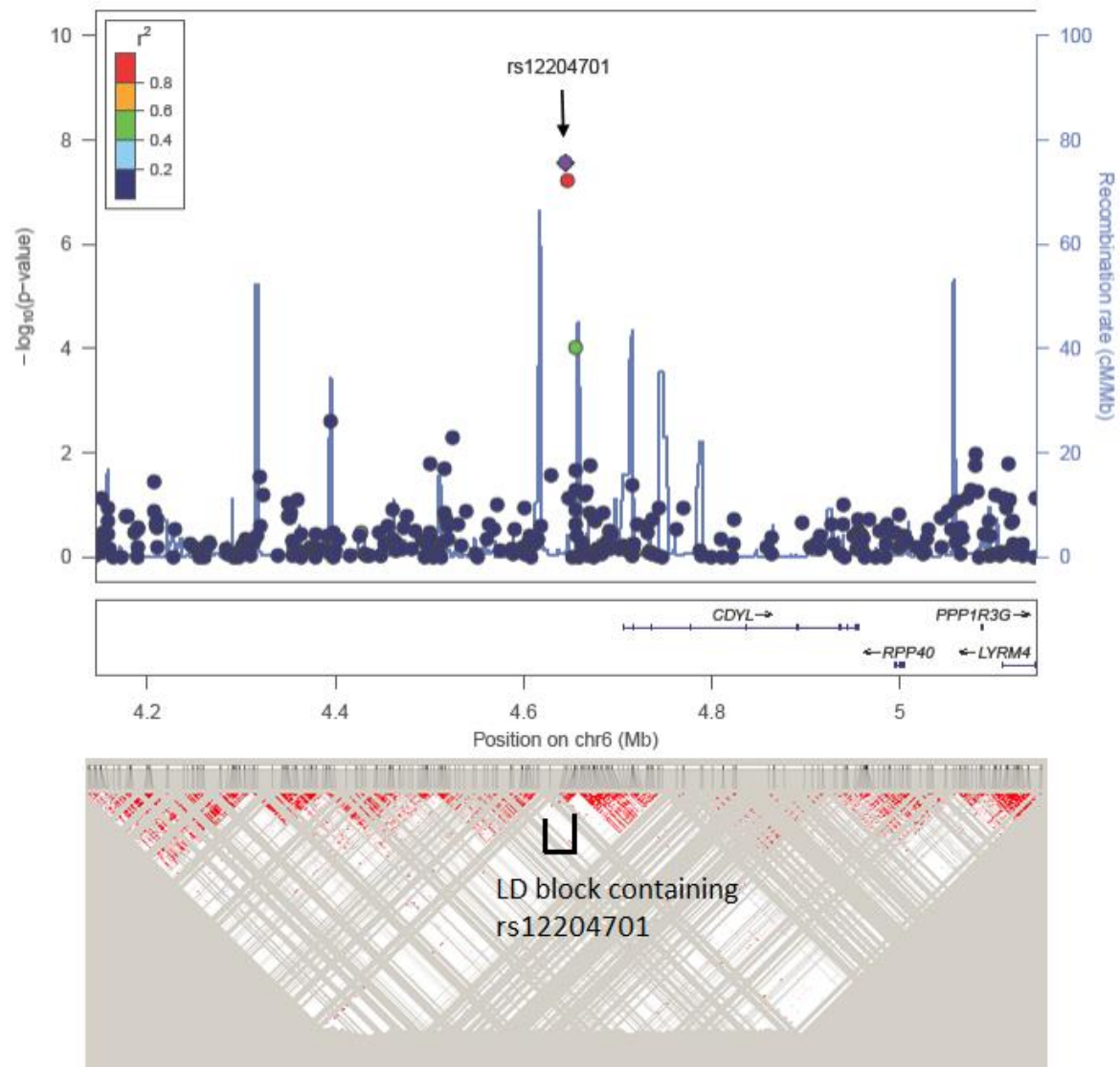


Figure 5.9: Regional association plot and linkage disequilibrium (LD) map for rs12204701 +/- 500Kb. In the association plot, the left y-axis represents $-\log_{10}(\text{P-value})$ for association with 3-month KD response; the right y-axis represents the recombination rate; the x-axis represents base-pair positions along the chromosome (human genome build 37). The top variant, rs12204701, is shown in purple, the rest of the variants are coloured according to their linkage disequilibrium r^2 value with rs12204701. In the LD map, LD is indicated as D'/LOD . The colour white = $\text{LOD} < 2$, $D' < 1$; blue = $\text{LOD} < 2$, $D' = 1$; shades of pink/red = $\text{LOD} \geq 2$, $D' < 1$; bright red = $\text{LOD} \geq 2$, $D' = 1$. LD pattern is based on genotype data obtained for this study. Confidence interval minima for strong LD: lower: 0.7, upper 0.98; Upper confidence interval maximum for strong recombination: 0.9; Fraction of strong LD in informative comparisons is at least 0.95; markers with < 0.05 MAF are excluded.

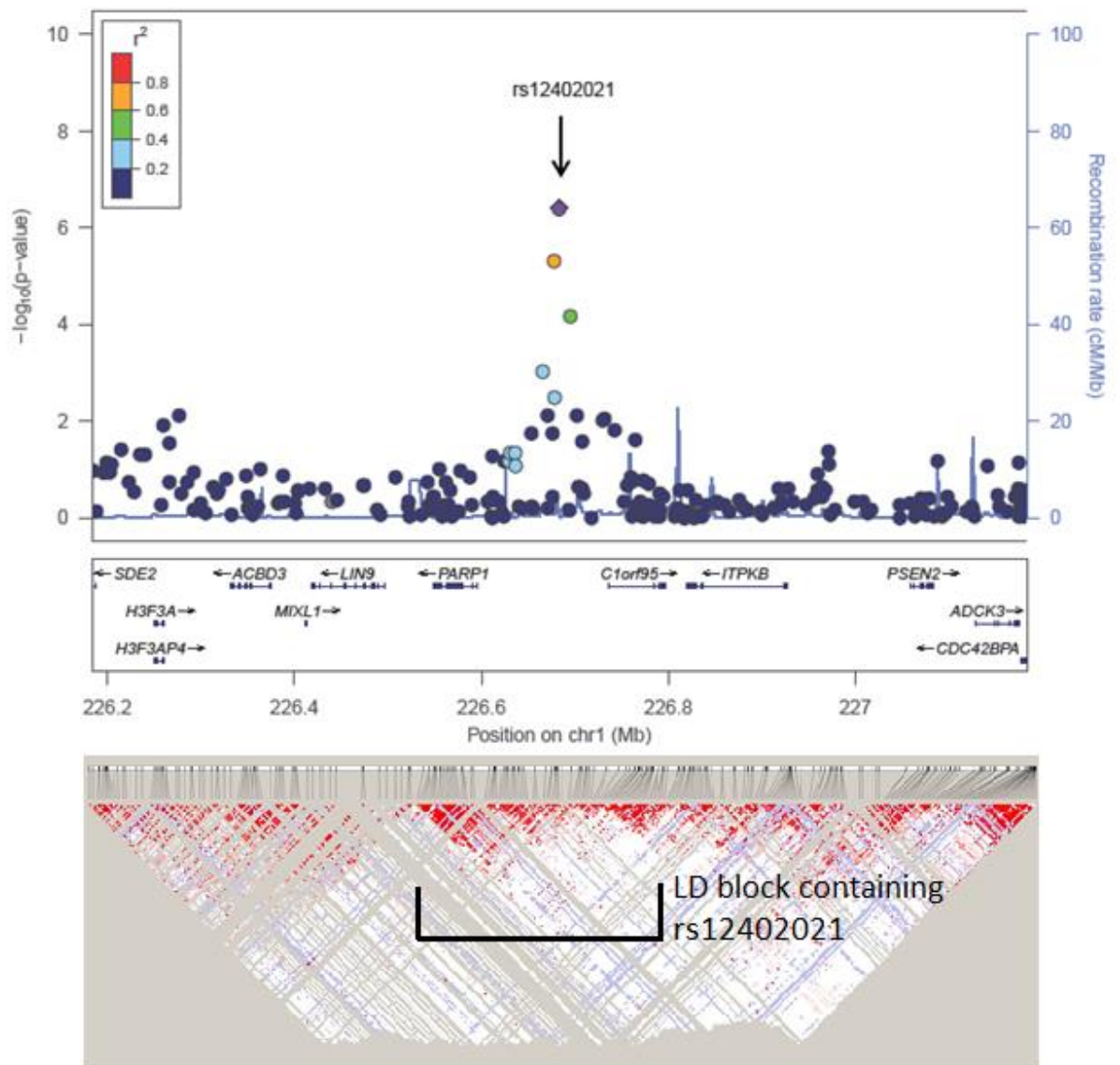


Figure 5.10: Regional association plot and linkage disequilibrium (LD) map for rs12402021+/- 500Kb. In the association plot, the left y-axis represents $-\log_{10}$ (P-values) for association with 3-month KD response; the right y-axis represents the recombination rate; the x-axis represents base-pair positions along the chromosome (human genome build 37). The top variant, rs12402021, is shown in purple, the rest of the variants are coloured according to their linkage disequilibrium r^2 value with rs12402021. In the LD map, LD is indicated as D'/LOD . The colour white = $LOD < 2$, $D' < 1$; blue = $LOD < 2$, $D' = 1$; shades of pink/red = $LOD \geq 2$, $D' < 1$; bright red = $LOD \geq 2$, $D' = 1$. LD pattern is based on genotype data obtained for this study. Confidence interval minima for strong LD: lower: 0.7, upper: 0.98; Upper confidence interval maximum for strong recombination: 0.9; Fraction of strong LD in informative comparisons is at least 0.95; markers with <0.05 MAF are excluded.

As shown in Figure 5.9, rs12204701 is in an LD block next to, but not within, the *CDYL* gene.

It is located in a region rich with recombination hot spots (by comparison, the mean sex-

averaged recombination rate in humans is 1.13cM/Mb [59]) and the LD structure is broken down because of this. Due to its ubiquitous expression in humans, *CDYL* is thought to be a potential housekeeping gene. Of the 22 proteins found to interact with *CDYL*, most play a role in transcriptional repression [703]. rs12204701 is also located 508770 bases from *ECI2*, which encodes a mitochondrial enzyme involved in the beta-oxidation of unsaturated fatty acids.

As shown in Figure 5.10, rs12402021 is located within an LD block with *PARP1* and *C1orf95*. *PARP1* encodes an enzyme that modifies various nuclear proteins by polyADP-ribosylation. It is involved in the repair of single-stranded DNA breaks and apoptosis and is thought to participate in the pathophysiology of Type I diabetes [704, 705]. *C1orf95* is uncharacterised.

rs4696766 is located in the intronic region of *ABLIM2*. This gene has not been extensively studied but it is associated with the GO annotations 'actin binding' and 'zinc ion binding' and it is assumed that the *ABLIM2* protein is necessary for the normal functioning of neurons [706].

rs4674639 is located in the intronic region of *PAX3*, a gene which encodes a transcription factor that is thought to regulate cell proliferation, migration and apoptosis. It is expressed early on in progenitors of the dorsal spinal cord and is thus thought to be involved in neural development [707].

As shown in Figure 5.11, rs12204701 is located adjacent to (but not quite overlapping) the transcript *KU-MEL-3*. Such long non-coding transcripts have been found to play a role in various cellular processes, including regulation of transcription and gene expression [708], and so variants located inside or adjacent to transcribed regions may be of interest.

Although this gene did not appear on the Locuszoom plot (presumably because it is an RNA

gene – the sequence of DNA from which a non-coding RNA is transcribed), its position at 6:4610846-4612153 indicates that rs12204701 is in high LD with it.

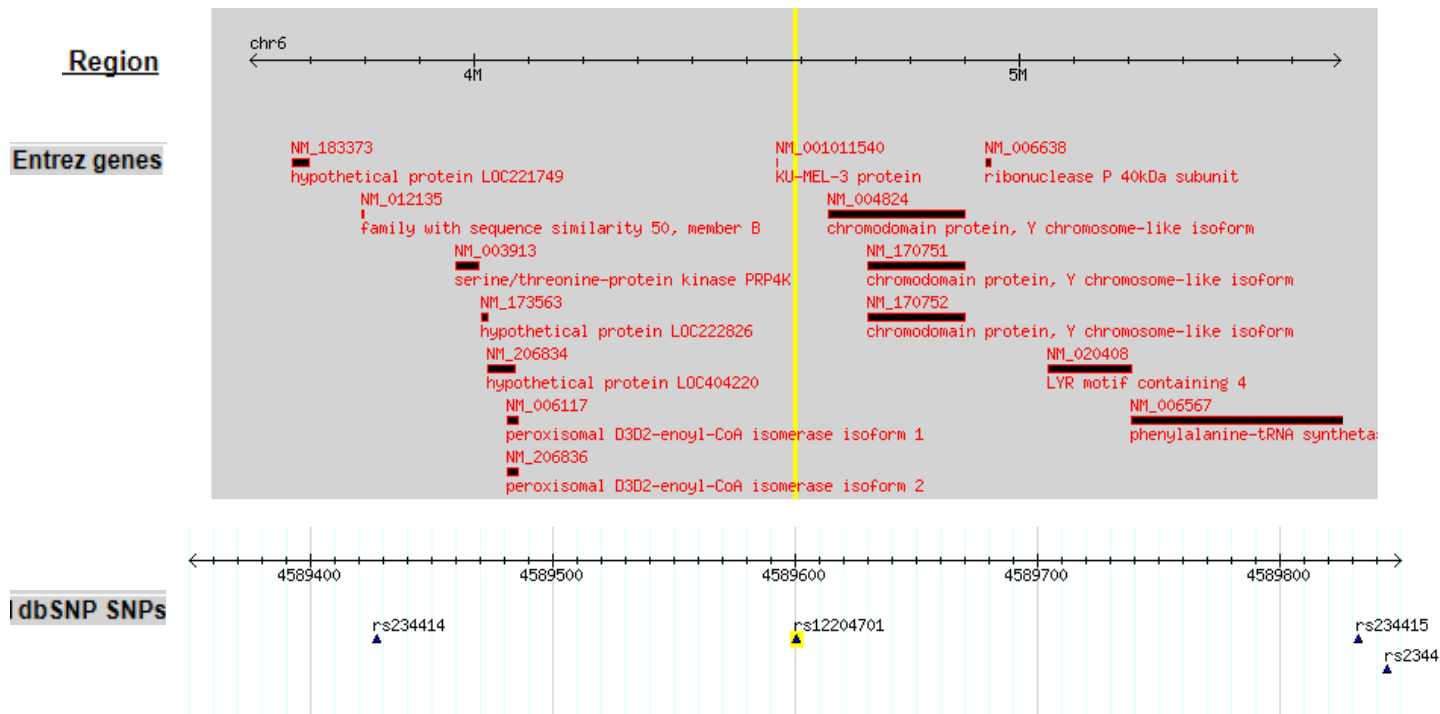


Figure 5.11: HapMap figure of rs12204701 +/-500Kb, including nearby Entrez genes and dbSNP SNPs

5.3.5 Power calculations

As shown by Figure 5.12, with the sample size used in the 3-month diet response GWAS

(238 subjects), only common variants (marker allele frequency approximately 0.1 to 0.2)

with a relative risk of approximately 3-4 could be detected with 80% power, assuming a co-

dominant penetrance model; on the other end of the spectrum, assuming a recessive

penetrance model, the marker allele frequency would have to be at least 0.45 to detect a

variant with a relative risk of around 6.

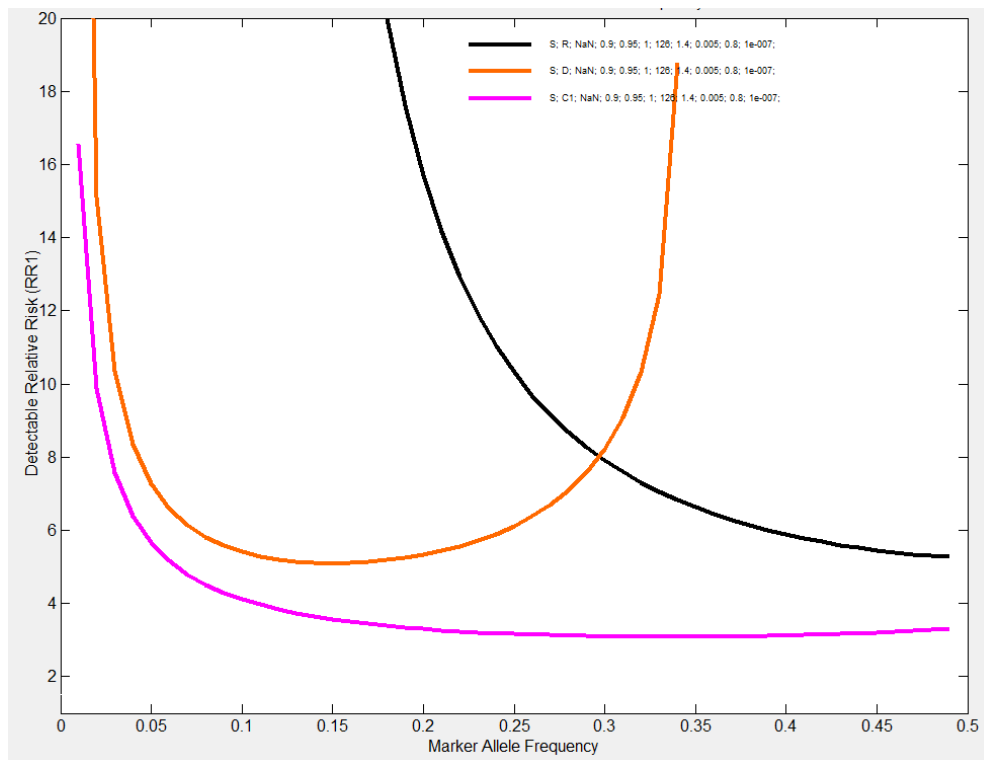


Figure 5.12: Detectable relative risk and disease allele frequency curves for 3-month KD response cohort, with 80% power, assuming r^2 of 0.9 between genotyped marker and causal variant, a disease prevalence of 0.5%, $\alpha=1.00 \times 10^{-7}$, 116 cases and control:case ratio of 1.01. Black line=recessive model, orange line=dominant model, purple line=co-dominant model

As shown in Figure 5.13, assuming a co-dominant or dominant genetic model, a sample size of approximately 2000 would have 80% power to detect variants with allele frequency 0.2 and relative risk of 1.4; a much larger sample size of approximately 14000 would be needed to detect variants with the same allele frequency and relative risk, assuming a recessive genetic model. Larger sample sizes would be needed to detect variants of lower allele frequency or lower relative risk.

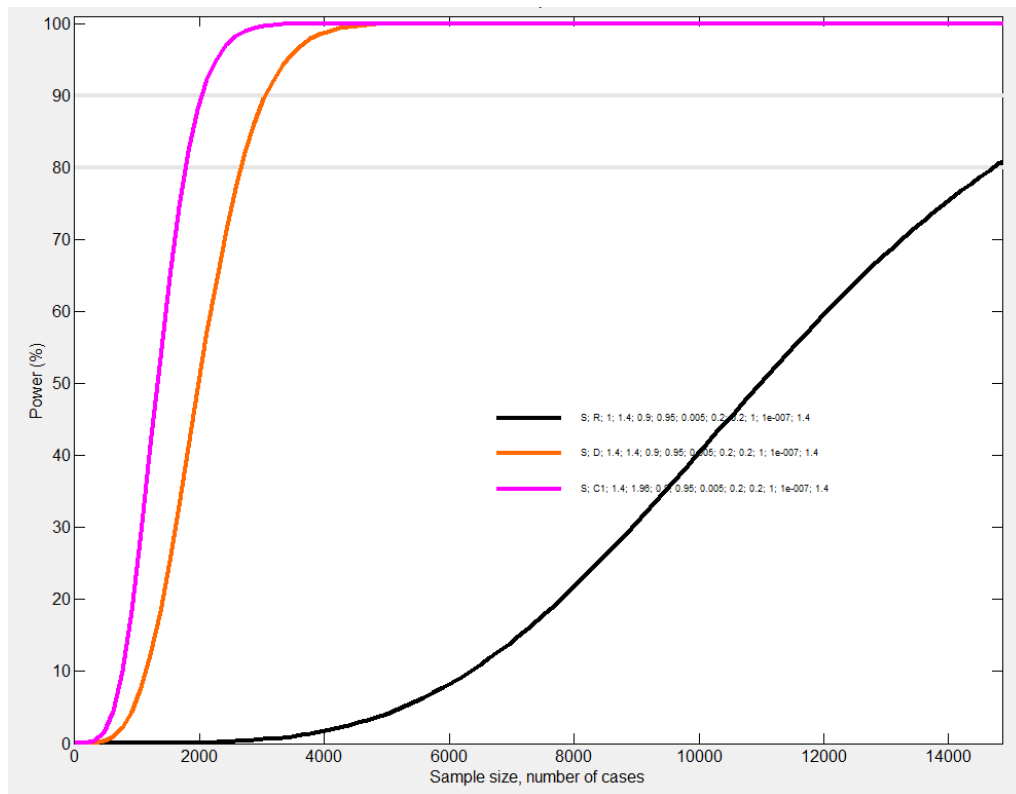


Figure 5.13: Power and sample size curves, using a recessive model (black line), dominant model (orange line) and co-dominant model (purple line), assuming $r^2=0.9$ between genotyped marker and causal variant, a disease prevalence of 0.5%, disease and marker allele frequencies of 0.2 and relative risk of 1.4, $\alpha=1.00 \times 10^{-7}$ and a control to case ratio of 1.4

5.3.6 Gene- and pathway- based analyses

A significance threshold of 2.59×10^{-6} ($0.05/19291$) and a suggestive significance threshold of 5.18×10^{-5} ($1/19291$) was set for the gene-based tests. A significance threshold of 3.79×10^{-5} ($0.05/1320$) and a suggestive significance threshold of 0.00076 ($1/1320$) was set for the pathway-based tests.

No gene or pathway reached significance or suggestive significance either in the 3-month or summary KD response analyses.

Nine genes had the lowest p-value ($p=0.000999$) in the 3-month KD response analysis and 12 genes in the summary KD response analysis ($p=0.000999$). These genes are given in Table 5.1, along with their associated GO terms and pathways, as identified from the online DAVID database and Gene Cards (www.genecards.org).

Table 5.1: Genes with lowest p-values in 3-month and summary KD response gene-set analysis

Gene abbreviation	Gene name	GO Biological Process	GO Molecular Function	Associated Pathways
Genes with lowest p-value in 3-month KD response analysis				
ARRB2	arrestin, beta 2	endocytosis, receptor-mediated endocytosis, cell surface receptor linked signal transduction, enzyme linked receptor protein signalling pathway, transmembrane receptor protein serine/threonine kinase signalling pathway, transforming growth factor beta receptor signalling pathway, sensory perception, membrane invagination, membrane organization, vesicle-mediated transport, receptor internalization, receptor metabolic process, neurological system process, cognition	KEGG pathways: MAPK signalling pathway, Chemokine signalling pathway, Endocytosis, Olfactory transduction, PANTHER pathways: Heterotrimeric G-protein signalling pathway-Gi alpha and Gs alpha mediated pathway, Inflammation mediated by chemokine and cytokine signalling pathway, Wnt signalling pathway, Angiotensin II-stimulated signalling through G proteins and beta-arrestin	-
C10orf68	chromosome 10 open reading frame 68	Not found Associated with Alzheimer's disease (http://www.genecards.org)		
C19orf35	clathrin, heavy chain-like 1	Not found Uncharacterised protein		
CLTCL1		intracellular protein transport, endocytosis, receptor-mediated endocytosis, protein localization, membrane invagination, protein transport, membrane organization, vesicle-mediated transport, cellular protein localization, establishment of protein localization, intracellular transport, cellular macromolecule localization	structural molecule activity	KEGG pathways: Lysosome, Endocytosis, Huntington's disease.

				<p>PANTHER pathways: Heterotrimeric G-protein signaling pathway-Gi alpha and Gs alpha mediated pathway, Heterotrimeric G-protein signaling pathway-Gq alpha and Go alpha mediated pathway.</p> <p>REACTOME pathways: Gap junction trafficking and regulation,</p>
FCN1	ficolin (collagen/fibrinogen domain containing) 1	immune effector process, ion transport, phosphate transport, anion transport, endocytosis, phagocytosis, immune response, opsonization, membrane invagination, inorganic anion transport, membrane organization, vesicle-mediated transport	antigen binding, calcium ion binding, sugar binding, carbohydrate binding, ion binding, cation binding, metal ion binding	-
MPO	myeloperoxidase	response to reactive oxygen species, oxygen and reactive oxygen species metabolic process, antiapoptosis, defense response, response to oxidative stress, response to inorganic substance, regulation of cell death, cellular response to stress, macromolecular complex remodeling, protein-lipid complex remodeling, plasma lipoprotein particle remodeling, low-density lipoprotein particle remodeling, cellular response to oxidative stress, cellular response to reactive oxygen species, response to hydrogen peroxide, hydrogen peroxide metabolic process, hydrogen peroxide catabolic process, regulation of apoptosis, negative regulation of apoptosis, regulation of programmed cell death, negative regulation of programmed cell death, macromolecular complex subunit organization, oxidation reduction, negative regulation of cell death, cellular response to hydrogen peroxide	chromatin binding, peroxidase activity, iron ion binding, calcium ion binding, antioxidant activity, oxidoreductase activity, acting on peroxide as acceptor, heme binding, ion binding, cation binding, metal ion binding, tetrapyrrole binding, transition metal ion binding	

RAB20	RAB20, member RAS oncogene family	intracellular signalling cascade, small GTPase mediated signal transduction, protein localization, protein transport, establishment of protein localization	nucleotide binding, GTP binding, purine nucleotide binding, guanyl nucleotide binding, ribonucleotide binding, purine ribonucleotide binding, guanyl ribonucleotide binding	-
SLC12A8	solute carrier family 12 (potassium/chloride transporters), member 8	ion transport, cation transport, potassium ion transport, monovalent inorganic cation transport, metal ion transport, transmembrane transport	symporter activity, potassium ion binding, chloride ion binding, alkali metal ion binding, ion binding, anion binding, cation binding, metal ion binding	-
XPO1	exportin 1	protein import into nucleus, docking, RNA localization, protein complex assembly, protein targeting, protein import into nucleus, protein export from nucleus, intracellular protein transport, nucleocytoplasmic transport, protein localization, response to virus, regulation of catabolic process, regulation of cell cycle process, regulation of centrosome duplication, protein transport, nucleobase, nucleoside, nucleotide and nucleic acid transport, protein import, regulation of microtubule-based process, regulation of organelle organization, protein localization in organelle, protein localization in nucleus, cellular protein localization, cellular macromolecular complex subunit organization, cellular macromolecular complex assembly, regulation of protein catabolic process, cellular protein complex assembly, macromolecular complex subunit organization, establishment of protein localization, regulation of centrosome cycle, intracellular transport, nucleic acid transport, RNA transport, mRNA transport, nuclear export, nuclear transport, nuclear import, establishment of RNA localization, regulation of cytoskeleton organization, regulation of cell cycle, macromolecular complex assembly, protein complex biogenesis, regulation of microtubule cytoskeleton organization, cellular macromolecule localization	RNA binding, protein transporter activity	REACTOME pathways: Cell Cycle, Mitotic, Influenza Infection, HIV Infection
Genes with lowest p-value in summary KD response analysis				
C10orf68	chromosome 10 open	Not found Associated with Alzheimer's disease (http://www.genecards.org)		

	reading frame 68			
CAMTA2	calmodulin binding transcription activator 2	muscle system process, transcription, regulation of transcription, DNA-dependent, regulation of transcription from RNA polymerase II promoter, positive regulation of biosynthetic process, positive regulation of macromolecule biosynthetic process, positive regulation of macromolecule metabolic process, positive regulation of gene expression, cardiac muscle adaptation, striated muscle adaptation, muscle hypertrophy, striated muscle hypertrophy, cardiac muscle hypertrophy, positive regulation of cellular biosynthetic process, muscle adaptation, regulation of transcription, positive regulation of transcription, DNA-dependent, positive regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process, positive regulation of transcription, positive regulation of transcription from RNA polymerase II promoter, positive regulation of nitrogen compound metabolic process, regulation of RNA metabolic process, positive regulation of RNA metabolic process	chromatin binding, calmodulin binding, transcription factor binding, enzyme binding, transcription regulator activity, histone deacetylase binding	-
CHRNA2	cholinergic receptor, nicotinic, gamma	muscle system process, ion transport, cellular ion homeostasis, muscle contraction, cellular homeostasis, regulation of membrane potential, homeostatic process, chemical homeostasis, ion homeostasis, cellular chemical homeostasis	ion channel activity, extracellular ligand-gated ion channel activity, cation channel activity, channel activity, ligand-gated ion channel activity, acetylcholine receptor activity, passive transmembrane transporter activity, ligand-gated channel activity, gated channel activity, substrate specific channel activity, neurotransmitter receptor activity, neurotransmitter binding, acetylcholine binding, amine binding, metal ion transmembrane transporter activity	PANTHER pathway: Nicotinic acetylcholine receptor signaling pathway
ENO3	enolase 3	monosaccharide metabolic process, glucose metabolic process, glucose catabolic	magnesium ion binding,	KEGG pathways:

	(beta, muscle)	process, generation of precursor metabolites and energy, glycolysis, aging, response to wounding, carbohydrate catabolic process, hexose metabolic process, hexose catabolic process, regeneration, growth, wound healing, tissue regeneration, response to drug, skeletal muscle regeneration, cellular carbohydrate catabolic process, alcohol catabolic process, monosaccharide catabolic process, developmental growth	phosphopyruvate hydratase activity, hydro-lyase activity, identical protein binding, protein homodimerization activity, ion binding, cation binding, metal ion binding, protein heterodimerization activity, protein dimerization activity	Glycolysis / Gluconeogenesis, RNA degradation. PANTHER pathways :Glycolysis, REACTOME pathways: Integration of energy pathways, Metabolism of carbohydrates, Diabetes pathways
HHLA2	HERV-H LTR-associating 2	No associated Biological Process or Molecular Function Encodes a protein that is thought to regulate cell-mediated immunity by binding to a receptor on T lymphocytes and inhibiting the proliferation of these cells (http://www.genecards.org)	-	
INCA1	inhibitor of CDK, cyclin A1 interacting protein 1	negative regulation of cell proliferation, negative regulation of cyclin-dependent protein serine/threonine kinase activity, positive regulation of apoptotic process	cyclin-dependent protein serine/threonine kinase inhibitor activity, cyclin binding, protein binding	-
LOC100652846	uncharacterised LOC100652846	Not found		
OR10J5	olfactory receptor, family 10, subfamily J, member 5	cell surface receptor linked signal transduction, G-protein coupled receptor protein signaling pathway, sensory perception, sensory perception of chemical stimulus, sensory perception of smell, neurological system process, cognition	olfactory receptor activity	REACTOME pathways: Signalling by GPCR. KEGG pathways: Olfactory transduction
PFN1	profilin 1	embryonic epithelial tube formation, neural tube formation, neural tube closure,	actin binding, cytoskeletal	KEGG pathways:

		morphogenesis of an epithelium, regulation of transcription, DNA-dependent, regulation of transcription from RNA polymerase II promoter, cytoskeleton organization, embryonic development ending in birth or egg hatching, primary neural tube formation, morphogenesis of embryonic epithelium, neural tube development, actin filament-based process, actin cytoskeleton organization, tube lumen formation, tube morphogenesis, tube development, chordate embryonic development, regulation of transcription, embryonic morphogenesis, tissue morphogenesis, regulation of RNA metabolic process, epithelium development, epithelial tube morphogenesis, tube closure	protein binding, Ras GTPase binding, Rho GTPase binding, enzyme binding, small GTPase binding, GTPase binding, proline-rich region binding	Regulation of actin cytoskeleton, PANTHER pathways: Cytoskeletal regulation by Rho GTPase, REACTOME pathways: Hemostasis
REPS2	RALBP1 associated Eps domain containing 2	protein complex assembly, cell surface receptor linked signal transduction, enzyme linked receptor protein signalling pathway, transmembrane receptor protein tyrosine kinase signalling pathway, epidermal growth factor receptor signalling pathway, macromolecular complex subunit organization, macromolecular complex assembly, protein complex biogenesis	calcium ion binding, ion binding, cation binding, metal ion binding	-
SHFM1	split hand/foot malformation (ectrodactyly) type 1	double-strand break repair via homologous recombination, recombinational repair, DNA metabolic process, DNA repair, double-strand break repair, DNA recombination, proteolysis, response to DNA damage stimulus, cellular response to stress	peptidase activity	KEGG pathways: Proteasome, Homologous recombination
ZNF273	zinc finger protein 273	transcription, regulation of transcription, DNA-dependent, regulation of transcription, regulation of RNA metabolic process	DNA binding, zinc ion binding, ion binding, cation binding, metal ion binding, transition metal ion binding	-

The PID 'PRL (protein of regenerating liver) Signalling events' pathway generated the lowest p-value in the 3-month KD response analysis ($p=0.0019$) and the summary KD response analysis ($p=0.00099$).

The PID PRL Signalling events pathway is within the pathway category 'Tyrosine-specific phosphatase mediated signalling pathways'. It contains the following genes: *AGT*, *ATF5*, *BCAR1*, *CCNA2*, *CCNE1*, *CDK2*, *CDKN1A*, *EGR1*, *ITGA1*, *ITGB1*, *MAPK1*, *MAPK3*, *PTP4A1*, *PTP4A2*, *PTP4A3*, *RABGGTA*, *RABGGTB*, *RAC1*, *RHOA*, *RHOC*, *ROCK1*, *SRC* and *TUBA1B*. A snapshot of the pathway (the whole pathway is very large) is given in Figure 5.14.

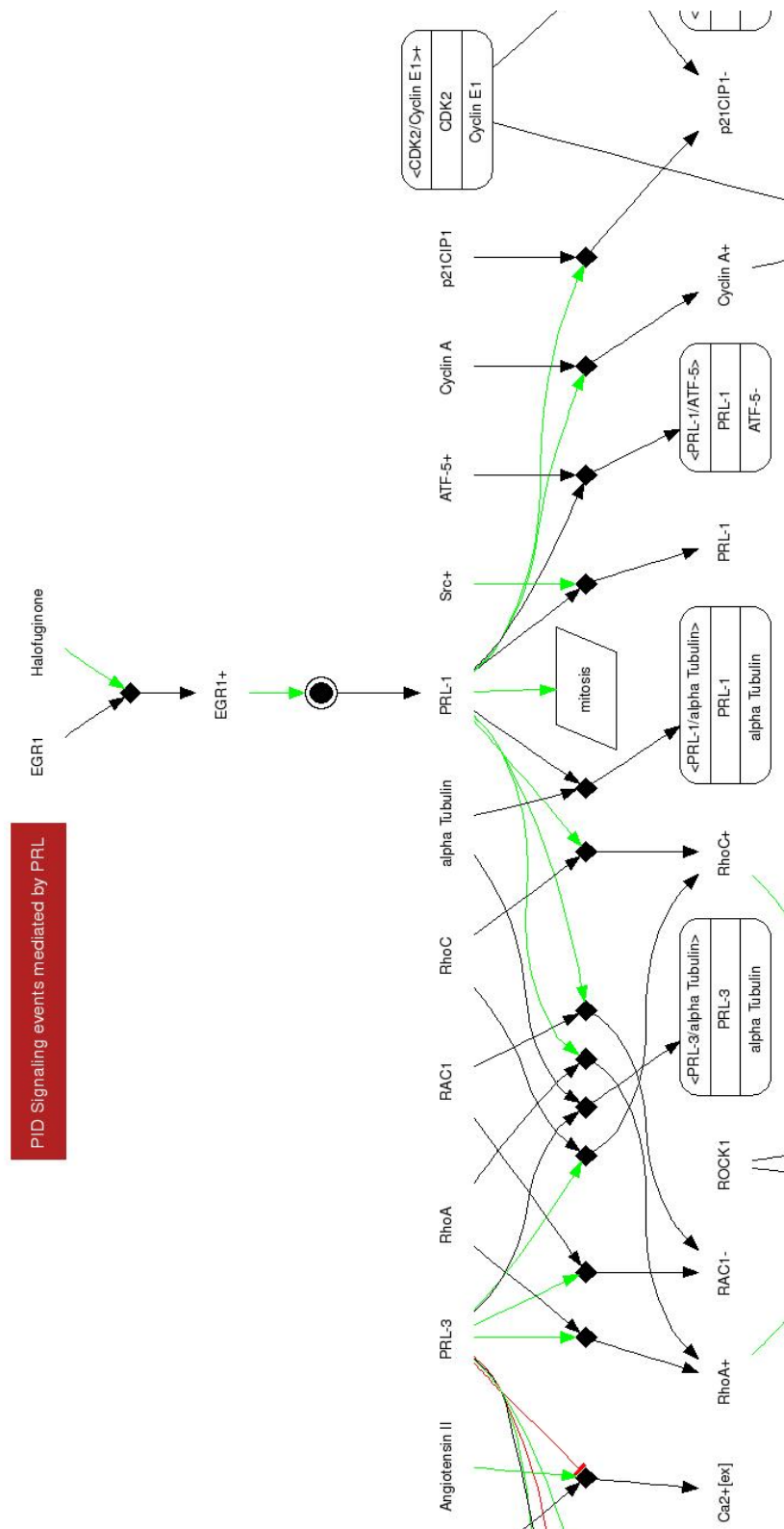


Figure 5.14: A Snapshot of the Protein Interaction Database PRL Signalling events pathway, taken from <http://pid.nci.nih.gov>.

5.4 Discussion

Key findings:

- The A allele in rs12204701 was significantly associated with unfavourable KD response in the 3-month diet response Fisher's exact test. This result is difficult to interpret.
- No significant results were obtained from the summary KD response single-SNP GWAS. This was most likely due to a lack of power to detect genotypic-phenotypic associations, even from common variants, due to the small sample size.
- SNPs of suggestive significance were within or in LD with genes that have a biologically plausible function with regards KD response.
- Gene and pathway-based analyses have highlighted some areas of interest, in particular regulation of cell survival/apoptosis.
- Different genes may affect KD response in the short- and long-term, although these genes may play similar roles and participate in the same pathways.
- Replication is needed with a larger sample size.

5.4.1 Significant results

The difficulty with interpreting the importance of the association of rs12204701 with 3-month KD response is that this SNP is located in a non-coding region and has no obvious function. Only 7% of SNPs previously found to be associated with a disease or trait in GWAS are located in protein-coding regions [649, 650]. Non-coding SNPs may alter expression of genes, rather than altering the protein-coding sequence or protein structure, presumably via transcription factor binding (variants that affect the ability of a transcription factor to bind to DNA) or by influencing gene splicing [709]. It is of interest that rs12204701 is located in the vicinity of, and is in strong LD with the lincRNA *KU-MEL-3* - a non-coding

transcript of >200 nucleotides length, which has been found to regulate gene expression [710, 711]. An association analysis of rs12204701 with (potentially tissue-specific) expression levels of *KU-MEL-3* would confirm any effect of these SNPs on the gene, as has been recently done with a range of lincRNAs [710].

rs12204701 is located close to *CDYL* in a region rich with recombination hot spots, where the LD structure is broken down. Furthermore, as a potential housekeeping gene that interacts with many different proteins, it is difficult to determine exactly how *CDYL* may influence KD response. It is of interest that *CDYL* is involved in the repression of transcription of genes, such as the proto-oncogene *TrkC*, which is important for suppression of cellular transformation [703]. Also, nearby SNPs, located between *KU-MEL-3* and *CDYL*, have been associated with phenotypic traits relevant to metabolism of high-fat, low-carbohydrate diet: cholesterol levels [712] and susceptibility to Type II diabetes [713]. rs12204701 may tag other SNPs or even copy number variants that may influence KD response. It may, of course, be the case that rs12204701 does not regulate its most nearby gene; for example, expression of genes, such as *GLI3* and *PAX6*, are at least partly regulated by *cis* elements distant from the gene itself [717].

The lack of significant results in the summary KD response GWAS is likely because of the lack of power – the sample size was even smaller than in the 3-month response GWAS. Most associations discovered from GWAS have effect sizes of 1.1–1.4 [714] and large sample sizes are required to detect associations with an odds ratio of 1.5–2 [715]. It may also be the case that causal/influential variants or genes are not in high LD with, or adjacent to genotyped variants, which further reduces the power to detect an association [716].

Another alternative is that causal/influential variants, despite being in LD with genotyped variants, have such small effect sizes that they do not reach statistical significance. These are most likely to be intermediate/rare variants [716], which will be explored in Chapter 6. Variants with small effect sizes may not have immediate implications for clinical practice, but may further understanding of the mechanisms behind the antiepileptic effects of the KD.

Other elements of GWAS methodology, such as the fact that SNPs are tested in isolation and that individuals have different environmental pressures acting on them, mean that associations may be missed if they are dependent on gene-gene or gene-environment interactions. Grouping SNPs together into sets may be a powerful way of detecting associations that would not otherwise be detected; by itself, an individual SNP may have a small effect on the phenotype but, in combination with other SNPs (which may all be in LD with the causal/influential variant), it may have a larger effect [718]. In this study, gene- and pathway-based approaches were also underpowered, as they were based on single SNP results. In combination with the single SNP approach, however, they did highlight some biologically plausible themes of interest.

The fact that different SNPs had the lowest p-values in the two GWAS (for example, rs12204701 did not even reach suggestive significance in the summary KD response GWAS) indicates that different variants/genes may play a greater role in influencing short-time KD response, whereas others come into play with more long-term response. Specific gene expression changes may materialise after different time periods following the onset of dietary treatment. It is not necessarily the case that taking KD response at a particular point in time is more or less accurate than the summary diet response phenotype (although, of course, an individual who achieves seizure reduction during the first three months on the KD may not maintain this seizure control during the subsequent three months) – they are

different models, both of which may help with understanding different aspects of the relationship (if any) between genetics and KD response. If a sufficiently large cohort was recruited (>2000 participants, according to power analyses), with long-term follow-up for as many patients as possible, appropriately-powered GWAS could be conducted using KD response at 3-, 6-, 12-, 18- and 24-month+ follow-up, in addition to a summary diet response model. This would allow investigators to better determine whether different variants/genes are relevant for KD response in the short- and long-term and, if so, at which point specific variants/genes cease to be relevant.

5.4.2 Potential sources of error

5.4.2.1 Phenotypic

As outlined in Chapter 3, the classification of individuals into responders and non-responders is somewhat arbitrary. Ideally, there would be standardised phenotypic criteria for defining KD response, such as the McDonald criteria for multiple sclerosis [719], but in this study, a seizure frequency threshold that is viewed as meaningful in clinical practice was used. Individuals or their parent/guardian(s) may have different definitions of what is clinically-significant for them, but ‘personalised’ definitions of response would make inter-individual comparisons very difficult. Furthermore, unless prospective data are specifically collected for research purposes, one is limited by the data available: for example, the only quantitative measure of response recorded in the study sites was seizure frequency and, in most cases, even this measure was not recorded on a daily basis. The measurement error, potentially introduced when relying on clinic letters for information, could have been reduced by more widespread use of seizure diaries, perhaps enabling use of a quantitative phenotype. This would have increased power to detect associations [720], and may have had a more interpretable outcome.

As there is no consistent evidence for an effect of any clinical or demographic factors on KD response, either in the literature or in our cohort (as outlined in Chapter 3), none were included as covariates. The inclusion of known covariates in GWAS may reduce power for detecting associations, depending on the prevalence of the trait being studied [722]. It is, of course, possible that other unknown, non-genetic covariates have led to Type I or Type II error.

5.4.2.2 Genotypic

The Illumina chip provides high coverage across the genome and powerful tag SNPs. With the exception of replicate genotyping of the same subjects, standard measures were taken to avoid genotyping errors and systematic bias, which decrease the power for detecting associations. For example, exclusion of SNPs with significant deviation from HWE, as well as low quality genotype calls, aims to reduce Type I error [723]. Due to the lack of precise guidelines for cut-off values for quality control variables in GWAS, these could be considered somewhat arbitrary and they do vary between studies. Bearing in mind thresholds used in other epilepsy GWAS, as well as the small sample size in this study, the effects of each variable was considered with the aim of striking a balance between maintaining only high-quality data without removing genuine phenotype-genotype associations. Potential bias due to batch effects was also considered. Changes in the number of samples in each batch (sample size affects the error involved in genotype calling for SNPs, particularly those of low MAF) and batch composition (in terms of proportions of cases and controls, and samples from different ethnic groups) may affect genotype calls [724]; discordant results may arise from varying probe synthesis efficiency between batches [725]. DNA concentrations may also differ between batches, or depending on the technique used for DNA extraction (two techniques are used at GOSH - Autogen and FujiFilm - and another is used in Melbourne, which is based on the same technology as Autogen [FlexiGene]), but preparation of all DNA samples to a specific concentration should

have ensured homogeneity. Although not comparing FlexiGene and filtration technologies specifically, various DNA extraction techniques have been found to perform similarly in terms of quantity and quality of the final DNA product [726]. Another potential contributor to batch effects - different DNA sources - was not applicable in this study, as all DNA was extracted from capillary blood. MDS plots showed that, even if there were differences in the data (for example, due to recruitment site or processing batch), these effects were random with respect to responder/non-responders status.

It is well-cited that population stratification may lead to confounding in GWAS. In the context of the KD, controlling for population stratification may be particularly important, as there may be a selection advantage against starvation in particular ethnic groups – responsiveness of the pancreas to exogenous glucose load, minimising postprandial excretion of glucose into the urine and favouring the storage of calories as fat, due to repeated exposure to periods of food shortage throughout history [727]. This ‘thrifty gene’ hypothesis has been suggested as an explanation why aboriginal populations, such as Native Americans, now have increased levels of obesity and increased risk of Type II diabetes and atherosclerosis [728, 729]; the genetic propensity to store excess calories as fat may be advantageous in situations of feast or famine, but not in the face of increased food availability and a more sedentary lifestyle. This genetic propensity may be beneficial in the context of KD diet-feeding.

The fact that λ_{GC} was 1.00 in both GWAS models suggests absence of importance of population stratification in this cohort. The notion of GC inflation assumes that the loci have roughly equal mutation rates and equal deviation from HWE and that existing population structure has a uniform influence throughout the entire genome. This disregards the fact that the genetic architecture of an individual is often an admixture of various genetic backgrounds [624], as well as SNP-specific differences [639]. λ_{GC} is thought

to adequately correct for population structure if enough loci (≥ 500 markers) are used to estimate the correction factor [730]. The number of markers used in this study was well above 500. The quantile-quantile plots also showed almost no deviation from the null across the entire distribution, which indicates lack of undetected population stratification or any other systematic differences in the data that may confound association studies, such as batch effect.

Ideally, non-Caucasians would have been excluded from recruitment in the first place to aid the creation of a homogenous sample set. However, this would have been practically difficult and self-reported ethnicity may not always be accurate – individuals may not be aware of their own ethnic admixture.

Appropriate statistical correction for multiple testing was undertaken in this study to avoid Type I error: the gold-standard, permutation, as well as Moskvina's effective number of tests method.

The effects of environmental factors may be an issue, although only extreme circumstances would result in changes in allele frequency in either cases or controls [639]. The population groups in this study were not selected based on specific demographic factors (aside from adherence to KD treatment), which may skew allele frequencies due to other factors. Take, for example, the selection of smokers as cases in a study - where allele frequencies may be affected by selection for nicotine addiction [639].

5.4.3 SNPs of suggestive significance

The SNPs of suggestive significance are within genes or in the same LD block as genes with roles that could be relevant to KD response.

For example, it is biologically plausible that variants that affect *PARP1* function could influence KD response. *PARP1* has been implicated in the pathogenesis of diabetes mellitus

[731-733]; *PARP1* activation (percentage of PARP-positive endothelial nuclei) was higher in individuals with Type II diabetes and in those with a family history of diabetes compared with controls [734]. The protein has been associated with diabetes-related complications, including endothelial dysfunction [735] and neuropathy [736, 737]. Variants in *PARP1* have also been associated with melanoma [738], Alzheimer's disease susceptibility [739] and protection against Parkinson's disease [740] (although there is conflicting evidence for the latter [741]). Inhibition of *PARP1* has been shown to be neuroprotective in models of cerebral ischaemia [742-744], in vivo Parkinson's disease model [745] and soman-induced seizure-related brain damage [746]. Potential neuroprotective mechanisms include improved neuronal conductance, reduction of cell death and reduction of ATP depletion [747]. The *PARP1* protein regulates the production of inflammatory mediators, such as inducible nitric oxide synthase, potentially through interaction with NF κ B; *PARP1* is thought to be a coactivator of transcription factors that regulate such immune and inflammatory response genes. It is also a substrate of caspases, mediating apoptotic and necrotic cell death [748]. The role of *PARP1* in mediating the death of pancreatic β -cells may be the reason for its association with diabetes [749]. Both antiinflammatory effects and prevention of neuronal cell death are postulated mechanisms of action for the antiepileptic effects of the KD, as outlined in Chapter 1. It is feasible that variants affecting any of these *PARP1* functions may affect response to the KD. The fact that rs12402021 is supported by nearby SNPs in high LD provides further reassurance that this result is not an artifact. This is, however, debatable: if one SNP is associated with a trait by chance, then another SNP in LD with the first SNP may also be associated by chance.

Interpretation of the importance of *ABLIM2* with regards KD response is difficult, as little is known about its function. Highest expression of *ABLIM2* has been observed in the brain and eyes [706]; others have found it to be most highly expressed in skeletal muscle, with

slightly lower levels in the brain (mostly in the caudate/putamen and hippocampus) [750]. ABLIM2 binds strongly to F-actin and interacts with striated muscle activator of Rho signalling (STARS) in order to enhance STARS-dependent activation of serum response factor. It has been suggested that (in muscle), ABLIM2 is involved in cellular adaptation to biomechanical stress and (in the brain), it is involved in delivering signals toward the actin cytoskeleton. The *C. elegans* homologue of ABLIM proteins, Unc115, has been shown to participate in neuron guidance [751-753]; due to its tissue-specificity, it has been speculated that the protein plays the same role in humans [750].

PAX3 is required for normal development of the embryo and mutations in the gene are associated with Waardenburg syndrome, characterised by craniofacial and limb abnormalities, deafness and pigmentation of the skin, hair and iris [754, 755]. The gene encodes a transcription factor that plays a role in the development of the neural tube and peripheral nervous system, leading to cell proliferation in the brain and differentiation of specific cell types [707, 756, 757]. Bone morphogenetic protein [758], Wnt [759] and Sonic hedgehog signalling [760] have been shown to be involved in controlling *PAX3* expression; different factors may be involved as development progresses [707]. *PAX3* is located within a region that has been linked with juvenile myoclonic epilepsy [761] and expression of the gene is upregulated in high-grade glioma tissues compared to low-grade and normal brain tissues, and expression increases as the grade of the tumour increases and patient survival rate decreases [762]. It is unknown exactly how mutations in *PAX3* may influence KD response, but the fact that this gene is involved in brain development and implicated in brain abnormalities suggests that it is a biologically plausible candidate.

5.4.4 Gene- and pathway-based analyses

No statistically significant results were obtained from gene or pathway analyses and many genes generated the lowest p-value in both 3-month and summary KD response analyses. A

higher number of permutations could have been undertaken to obtain p-values that were accurate to a greater number of decimal places, but the computational cost of this would be large. Furthermore, it is likely that more than one gene is responsible for variability in KD response, and that some have greater effects on KD response, compared to other genes. Some common themes were present amongst the functions of the genes with the lowest p-values: cell cycling, neurological processes, immune/stress response, glucose metabolism, ion and cation binding/transport and GTP or GTPase binding/regulation. Most genes were different in the 3-month and summary KD response lists, although the same themes pervaded. The only gene that did overlap was *C10orf68*, which is of unknown function, but is located within a region that has been associated with late-onset Alzheimer disease [763].

The fact that many genes generated equally low p-values in the set-based tests, together with the fact that several common themes can be identified amongst their functions, indicates that investigation of pathways may be more fruitful than investigation of single genes. This section will therefore concentrate on the pathway with the lowest p-value in the 3-month and summary KD response analyses (PRL signalling events).

PRL proteins, which consist of PRL1, PRL2 and PRL3, are a protein tyrosine phosphatase subfamily, which are ROS effector proteins (they are reversibly oxidized by ROS) [764]. PRLs are thought to play a role in cell growth, proliferation, migration and invasion [765, 766], potentially by increasing expression of cyclins that are critical for cell cycle transitions [767]. PRL1 was found to be mainly localised in the brain and muscle, with increased expression in the cerebral cortex following ischaemia [768], PRL2 was found mainly in skeletal muscle, and PRL3 in cardiac and skeletal muscle [769]. PRLs are thought to promote tumorigenesis and metastasis, and overexpression of PRLs, in particular PRL3 and PRL1, has been frequently found in tumour tissues and metastatic human cancers [764]. When PRLs are in

their reduced state (not oxidised), their phosphatase activity is active and they may promote cancer metastasis; when oxidised by ROS, their phosphatase activity is inhibited and so metastasis is not promoted. It is unknown exactly how PRLs regulate metastasis, and they may exert their effects via various pathways [770]. PRL activation has been found to stimulate the PI3K/AKT pathway, which has an antiapoptotic effect [771], as well as stimulating phosphorylation of mTOR, which is mediated by PI3K/AKT (PI3K activates serine/threonine protein kinase AKT, which then activates mTOR), and regulates the start of translation and entry into the cell cycle [772]. PRL can activate the ERK1/2 pathway, which is involved in cell survival and proliferation [773], by interacting with integrin β 1. Others have reported that PRL3 exerts its effects via regulation of the Rho family GTPase [774, 775], activation of Src [776] or janus tyrosine kinases (JAKs) [777].

This pathway could feasibly contribute to the antiseizure effects of the KD and thus genetic variation that affects this pathway could account for variability in KD response. The biological functions highlighted in the PRL pathway (integrins/cell cycling/regulation of survival and apoptosis) have many implications. Through binding of integrins to signalling intermediaries such as SRC and PI3K, which in turn increases tyrosine phosphorylation and signalling through RAS, RAF, MEK, and small GTPases (such as Rho and Rac), a wide variety of basic cellular functions are influenced, including cell cycling, gene expression and cell survival [778]. There are a plethora of brain-specific functions that could be relevant in this cohort of people with drug-resistant epilepsy: for example, integrins are thought to be involved in stabilising long-term potentiation and synaptic function [778]; the MAPK pathway is known to play a role in survival of neurons [779, 780]; dysregulation of SRC family kinases has been said to contribute to epileptogenesis by regulating ion channel activity and synaptic transmission [781, 782]; cell cycle regulators have been found to be

necessary for neuronal cell death in response to injury and dysregulation of the cell cycle in neurons has also been associated with epilepsy [783].

Furthermore, these pathways have been shown to regulate insulin sensitivity and thus contribute to glucose homeostasis: the PI3K/AKT has been said to regulate cell survival/death via its effects on glucose uptake (promoting the translocation of glucose transporters to the cell membrane) and metabolism (stimulating activity of hexokinase, the first enzyme in the glycolytic pathway) [784, 785]; the MAPK family member ERK is needed for increased glucose uptake and glycolysis during T cell activation (this is for the survival, proliferation and cytokine production of activated T cells) [786]; SRC inhibition has been shown to decrease insulin-induced glucose uptake [787]; the Rho family GTPase RAC1 has been implicated in the regulation of insulin-dependent glucose uptake and shown to influence insulin-dependent translocation of GLUT4 to the cell membrane in skeletal muscle cells [788-790]. These results suggest that genetics may play a role in KD response, but they do not point to a single mechanism of action that is responsible for variability of KD response in this cohort.

Aside from the general nature of these categories, the results of the pathway analyses do not give any information on the direction of the effect on KD response. P-values are derived from the single-SNP analyses, which simply reflect a difference in frequency of variants between responders and non-responders. It may be that some genes in the PRL signalling events pathway have a favourable effect on KD response, whilst others have a detrimental effect.

Another point to consider when interpreting the results of gene- and pathway-based analyses is that, when using technology such as the Illumina chip, one is limited by the content of the array. Arrays aim to provide genome-wide coverage, but not coverage

within every gene. In these analyses, variants were only assigned to a gene if they fell within the gene or its transcript. The benefit of this is that one can be confident that these variants are associated with the 'correct' gene (if variants are present in an LD region in which there are many genes, this complicates assignment to a gene) and their functional consequence is easier to interpret. The disadvantage is that many potentially useful intergenic variants are excluded. One may use certain programmes to assign GWAS SNPs to genes, but these assignments are not certain and are often based on pre-existing knowledge about genes/variants associated with the phenotypic trait. Furthermore, interpretation of the functional consequences of such (predominantly non-coding) variants is often problematic. Use of sequencing data, in which (subject to error) every variant present is captured, eliminates this bias. Sequencing data could also be used for gene-based tests that assume a uni-directional effect of variants. This would help to determine the direction of effect of certain genes, which may lie within a certain pathway.

Different definitions and classifications of pathways may generate different results. For example, in a previous study, one over-represented pathway was the PANTHER pathway 'inflammation', which included 315 genes [702]. When using other pathway definitions, these 315 genes were divided up into smaller pathways that were related to inflammation. To avoid such bias, it is important to use a range of pathway definitions and to explore the top associated pathways to determine whether they can be linked together. Results also depend on the algorithm used, for example, whether the SNP with the lowest p-value is used as the set-wide p-value or an average of all SNP p-values in the set. In order to avoid potentially biased application of any one algorithm, one may use a variety of approaches and compare those genes or pathways that overlap [665]. The number of algorithms one uses is obviously time-dependent and could easily be adopted as a topic of focus by itself.

5.4.5 Conclusions and future work

rs12204701 was significantly associated with KD response at the 3-month point, but no significant results were obtained from the summary KD response single-SNP GWAS. These GWAS were underpowered to detect phenotypic-genotypic associations, even for variants of $MAF > 0.05$ and large effect sizes, and so it cannot be confirmed that other common variants do not influence KD response. Some variants with the lowest p-values are of interest due to their proximity (and LD patterns) to genes associated with metabolic and/or neurological processes. The fact that different results were obtained in the various GWAS suggests that different variants and genes may be relevant to KD response in the short-term and long-term. Cell cycling, regulation of cell death/survival and glucose metabolism may be relevant to both short- and long-term KD response, although different genes may play more of a role in different individuals or at different time points.

The logical next step would be a higher-powered study, achieved with a larger, ethnically-homogenous cohort. Although this would greatly increase the duration and cost of data collection, in this case, a lack of statistically significant associations could not be blamed on a lack of power. GWAS could be conducted with KD response classified at specific time points (if long-term follow-up data were available for a large enough cohort) or as a summary of response over time. If statistically significant associations were then observed, the most significant variants should be tested in an independent replication cohort, ideally using an independent genotyping platform [654].

'Missing' SNP genotypes (those not present on the genotyping array) may be extrapolated by imputation using LD patterns, for example, from HapMap. The benefits of imputation are that it is a cost-effective way of increasing coverage, and it increases power and allows pooling of data from various GWAS in meta-analyses [791]. Imputation provides a higher-resolution view of the region of interest, facilitating identification of casual SNPs [792].

Other sources of structural variation can also be imputed from genotyped SNP, such as CNVs [792] and common small insertions and deletions (indels) [793]. Imputed missing data has been shown not to significantly impact on the statistical properties of the data [794], although it must be considered that the ability to impute a genotype for a given SNP depends on the accuracy of the genotyped SNPs in the region and their correlation to the 'missing' SNP [791]. Particular care should be taken with rare variants, as the asymptotic theory may not hold and the effects of population structure may lead to false positives [792].

If a variant or gene were to be securely implicated with differential KD response (from GWAS with bigger sample sizes and replication of results in other cohorts), fine-mapping of trait-associated loci or animal models assessing the impact of knocking-down or deleting the region or genes in question, may aid demonstration of causality [795]. At the very least, evidence should be provided that the variants in question show a consistent association with factors that are highly correlated with KD response, such as regulation of expression of a gene associated with a metabolic or neurological phenotype.

6 Rare genetic variation and response to the Ketogenic diet: Illumina exome chip

6.1 Introduction

6.1.1 The common disease/rare variant hypothesis

An alternative to the common disease/common variant hypothesis, upon which most GWAS are based, is that, despite the fact that most genetic variation is common, most disease-causing variants are rare [716]. Based on evolutionary theory, if these variants have a deleterious effect on fitness, there will be a negative purifying selection pressure acting against them. Recent studies have reported that non-synonymous single nucleotide variants (SNVs) with lower MAF are more likely to be non-neutral (disease-related) than those with higher MAF [555, 588, 796]. This is not surprising, as rare variants are most often present in protein-coding regions [797, 798].

In support of the rare variant hypothesis is the fact that, even in the most successful GWAS, variants associated with complex diseases only tend to account for a small proportion of heritability [715]. This ‘missing heritability’ may be at least partly accounted for by low MAF or rare variants. These rare variants may not be tagged by common SNPs in GWAS as, due to their low MAF, they often are not in LD with common (or indeed other rare) SNPs [799]. Alternatively, it may be the case that other variants have not yet been found to be associated with disease because their effect sizes are too small to reach statistical significance, especially with single-SNP analyses and limited sample sizes [800]. There is increasing evidence that rare variants play a role in complex diseases: for example, five rare variants in *NOD2*, which act independently of one another, were found to be associated with Crohn’s disease risk [801-803]; rare variants that impair melatonin receptor 1B function were found to collectively contribute to Type II diabetes risk [804]; a rare missense

variant in *MYH6* was found to be associated with an approximately 12-fold increase risk of sick sinus syndrome [805].

An immediate difficulty that comes to mind when considering how to detect rare variant genotypic-phenotypic associations is the issue of sample size. The statistical power to detect an association is lower for rare variants than for common variants with a similar effect size, due to the small number of observations per variant and the higher likelihood of genotyping errors [796, 806]; one may thus assume that large sample sizes (larger than for common variants) are needed when looking at the effects of rare variants, particularly if the effect sizes of these variants are small [714]. It is expected, however, that rare variants have larger effect sizes than common variants, which may be detectable with smaller cohorts. As common variants tend to be older than rare variants, common variants (in order to remain common) must have survived purifying negative selection pressures over time; this is less likely to be the case if they have large deleterious functional effects, unless the variant is still in the process of being depleted from the population [807]. Many rare Mendelian and complex diseases are due to rare variants with large effect sizes [808], such as variants within *CFTR* for cystic fibrosis [809] and *BRCA1* and *BRCA2* for breast cancer [810].

In addition to variant-pooling approaches (discussed in Chapter 5), there are various approaches to increasing power in rare variant association studies. These are discussed below.

6.1.2 Variant prioritisation: variant class

A common approach for increasing power in genetic association studies is to incorporate information from functional prediction algorithms by selecting only functional variants – those that are more likely to be causal – for analysis [806, 807]. This further reduces the multiple-testing burden. If only putatively functional variants are included in analyses, one

has more confidence that they have a functional effect [807]. However, distinguishing between variants with and without functional consequence is not an easy task, as there is no standard classification system used to categorise variants by predicted functionality. For example, non-synonymous variants may disrupt biochemical function but not necessarily lead to disease or influence the phenotypic trait in question; it may depend on interactions with other variants or genes. It must also be considered that not all protein functions are known and so a variant may erroneously be classified as non-susceptible.

When trying to predict functionality of variants, a common first step is to divide them into variant classes according to their likelihood of altering the protein sequence or gene expression. Non-synonymous coding variants are primarily of interest in association studies: they lead to a substitution in the encoded amino acid and thus may alter the protein product, which may affect a phenotypic trait, and they have a lower MAF and occur less frequently in most protein-coding genes compared to synonymous variants [811], which indicates that they are under strong purifying selection [812]. Over half of the disease-associated variants in the Human Gene Mutation Database are non-synonymous [813]. 25–50% of rare (MAF<0.5%) non-synonymous variants are estimated to be deleterious [814]. Non-synonymous variants may be classified as missense (a variant that changes the encoded amino acid to another amino acid), stop-gain (a variant that changes a codon to a stop-codon), stop-loss (a variant that changes at least one base in a stop-codon), indels (insertion or deletion of a number of bases; can be in multiples of three [in frame or non-frameshift], or not in multiples of three [frameshift]), or splice-site (a variant located within the region of a splice site).

Loss-of-function variants, which have been defined as stop-gains, frameshift indels in coding sequences and disruptions to essential splice sites, have a high probability of being deleterious [815]. Stop-gains and frameshift indels that indirectly lead to a premature stop

codon may cause degradation of the entire transcript via nonsense-mediated decay – a quality control mechanism that targets and eliminates mRNAs that contain premature stop codons in order to reduce errors in gene expression and deleterious gain-of-function or dominant-negative protein products, where the gain-of-function is different from that of the original protein [816]. Both in highly-conserved coding regions and non-conserved regions, most stop-gain variants and, to a lesser extent, disruptions to splice sites, are rare (MAF<0.5%), indicating that these mutations can be highly deleterious, whatever the level of sequence conservation [814]. Coding variants that affect splicing, both within the splice junctions that define exon boundaries or those located near to these junctions, can lead to the creation of unstable mRNA and defective protein structure, or cause deleterious effects by altering the balance of expression of different protein isoforms produced from alternative splicing [817]. Variants may affect exonic splicing enhancers, which assist in splice-site recognition of exons and promote exon inclusion (gain of sequences), or exonic splicing silencers, which inhibit exon inclusion (loss of sequences). Annotation software can be useful for prioritising loss-of-function variants, as known stop-gain variants can be identified, both in exonic and exonic splicing regions. However, information may not be given about the effects of variants in non-coding regions, such as 3' and 5' UTRs, despite the fact that they may cause loss-of-function. For example, functional analyses revealed that the c.-21C>T mutation in the 5' UTR of *CDKN2A*, found in melanoma patients but not in controls, lead to a 'severe' negative impact on luciferase reporter activity, compared with the wild-type 5' UTR sequence, indicating that it predisposes individuals to melanoma [818].

Furthermore, loss-of-function variants are not all deleterious: they can also have small phenotypic effects, for example, if they disrupt the function of a non-essential gene or a pseudogene [819]. The fact that a typical 'healthy' person bears approximately 100 loss-of-

function variants in their genome, most of them in the heterozygote state [815], suggests that loss-of-function variants are often benign – stop-gain variants may lead to a truncated but still functional protein and functional transcripts can still be generated from alternative splicing; they may even be beneficial, as is the case with caspase-12 where a loss-of-function SNP increases sepsis resistance and decreases mortality risk in hospital surroundings [820]. Even alternative splicing variants that affect a large proportion of the transcript, potentially leading to exon skipping (a type of alternative splicing, where an exon [a ‘cassette’ exon] is spliced out of the transcript along with its flanking introns [821]), may in fact partially rescue the biological function of the gene. This has been shown with the dystrophin gene, where exon skipping can restore the open reading frame so a shortened but still functional dystrophin protein is produced, leading to the milder Becker muscular dystrophy instead of Duchenne muscular dystrophy [822]. The deleterious effect of frameshift indels can be limited by nearby compensatory indels that restore the reading frame, or those that occur near the 3'-end of genes that have no effect on gene sequence or length [823].

The functional consequences of other non-synonymous variants are also difficult to predict. Approximately 20% of rare missense mutations in humans result in a loss of function, 53% are subject to purifying selection, and around 27% are effectively neutral [824]. In-frame indels are often predicted to be of neutral impact, as the reading frame is maintained. They can, however, be strongly deleterious when located in exons, or when the indel does not coincide with codon boundaries [825]; selection against exonic in-frame indels is similar to that against missense variants [826]. Stop-loss variants, which result in an elongated transcript, may be harmless or render the protein non-functional.

Synonymous variants, which require different transfer RNAs to decode the single base pair changes but do not lead to a change in the encoded amino acid, tend to be considered

functionally neutral. However, as has been demonstrated, for example, with variants within *ADAMTS13* [591], synonymous variants can affect protein function and expression and are subject to evolutionary constraint and/or conservation [590].

Non-coding variants may also contribute to disease risk. About 10% of the non-coding genome is selectively constrained [579], suggesting that it is of some functional importance [590]. Non-coding variants may impact the genome through modulation of gene expression. This has been shown with a 20kb deletion, which is in high LD (r^2 of 1.0) with a SNP that is strongly associated with Crohn's disease (rs13361189), and which overlaps the promoter for the gene encoding immunity-related GTPase family M protein (*IRGM*), altering its expression [827]; the authors suggest that the association of Crohn's disease and *IRGM* arises from changes in *IRGM* gene regulation, which affects the efficacy of autophagy.

Intronic variants, including those within 5' and 3' UTRs, can act as intronic splicing enhancers and silencers - variants in upstream 3' splice sites tend to inhibit splicing and those in downstream 5' sites tend to enhance splicing [828]. Variants within these splicing regulatory elements (just like exonic splicing enhancers and silencers) can interfere with the recognition of splice sites, which is crucial for the correct excision of introns, and may cause exon skipping, which can affect the 3D structure of proteins or its interactions with other elements [829]. For example, a mutation within intron 7 of *SMN1* (survival motor neuron 1) was found in patients with spinal muscular atrophy [830]; this variant is predicted to disrupt the consensus exon 7 splice donor motif and it results in increased production of the alternatively spliced transcript *SMNΔ7* and low levels of full-length form SMN.

Variants affecting any of the following aspects of 5' UTRs can lead to impaired protein synthesis through altered regulation of translation: the length or secondary structure of 5' UTRs, the upstream open reading frames, internal ribosome entry site or iron-responsive elements present in 5' UTRs [831]. The length of the 5' UTR determines how much energy a ribosome needs to reach the AUG start codon, thus influencing translation efficiency. Variants that reduce the length of upstream open reading frames, that cause a frameshift in them, eliminate them or induce creation of novel upstream open reading frames, can also affect translation, either by destabilising the mRNA or by causing the ribosome to dissociate at a premature stop codon and potentially start re-scanning at a downstream open reading frame. Variants that affect internal ribosome entry sites can also influence regulation of translation and have been associated with various diseases, such as Charcot-Marie-Tooth disease and fragile X syndrome. Variants that affect the stability or structure of iron-responsive elements (regions of secondary structure within the 5' UTR to which mRNA-binding proteins bind in order to repress or promote translation) have also been associated with disease, such as hereditary hyperferritinaemia/cataract syndrome [831, 832].

Evidence regarding the functionality of 3' UTRs is not as strong as it is for 5' UTRs, but this region of mRNA has been shown to influence the translation, localisation and stability of mRNA [831]. Variants that influence the stop codon, either increasing or decreasing the length of 3' UTRs, can lead to premature termination of translation and formation of truncated polypeptides, or to elongated mutant proteins that may interfere with the functions of the normal protein or those proteins with which it associates. Variants in the polyadenylation signal – a highly-conserved sequence motif recognised by RNA-binding factors – although rare, can lead to premature termination of transcription and decrease the efficacy of cleavage and polyadenylation of pre-mRNA [833]. Variants that affect the

secondary structure of 3' UTRs (such as sequence alterations in the 3' UTR of *GATA4*, predicted to alter RNA folding and contribute to cardiac malformations and coronary heart disease risk [834]) alter its interaction with associated proteins and may affect translation efficacy of the mRNA transcript and transport to its site of function [835].

This demonstrates that, although non-synonymous variants and those within regulatory regions are more likely to affect gene expression and protein function compared to synonymous and non-coding variants, relying on variant class alone does not guarantee accurate prediction of functionality/neutrality. Additional information is needed to prioritise putative disease-associated variants.

6.1.3 Variant prioritisation: bioinformatics tools

Bioinformatics software, which is particularly useful when a large number of variants needs to be categorised, can also provide a variety of prediction tools based on molecular evolutionary patterns, protein sequence and/or structure. For example, the 'sorting tolerant from intolerant' (SIFT) algorithm is based on the premise that highly-conserved amino acids tend not to tolerate substitutions, whereas those with a lesser degree of conservation tolerate most substitutions [836]. The probability is calculated that an amino acid at a certain position is tolerated, conditional on the most frequent amino acid being tolerated. The 'Polymorphism Phenotyping 2' (PolyPhen 2) algorithm is, along with SIFT, the most widely used algorithm [837] and both have been found to be among the top most accurate methods for predicting functionality [838]. PolyPhen2 calculates the probability that a given mutation will be damaging, based on various sequence- and structure-based predictive features, as outlined in the algorithm release publication [576]. Other algorithms include: the 'likelihood ratio test' (LRT), which compares the null model - that each codon is evolving neutrally, with no difference in the rate of non-synonymous to synonymous substitution - to the alternative model - that the codon has evolved under negative

selection with a free parameter for the ratio of non-synonymous to synonymous substitution rates of the codon [577]; MutationTaster, which integrates information from various databases, such as dbSNP, the splice site prediction programme NNSplice, and SwissProt, and evaluates results with a naïve Bayes classifier [578]; ‘Genomic Evolutionary Rate Profiling++’ (GERP++), which identifies constrained elements by computing position-specific expected rates of evolution with use of a maximum likelihood estimation; and PhastCons [839], which is based on a phylogenetic hidden Markov model, a statistical model of sequence evolution.

The reliability of bioinformatics tools has been called into question, particularly considering that classifications from different methods may not always overlap [577, 588, 806]. The criteria generally used to evaluate these methods are false positive rate (the percentage of substitutions predicted to be damaging on substitutions known to be functionally neutral), false negative rate (the percentage of substitutions predicted to be functionally neutral on a set of substitutions that are known to affect protein function), and accuracy (the number of true positives plus negatives divided by the number of total positives plus negatives). Evaluation measures reported in various studies for some of the aforementioned algorithms are given below:

SIFT: 20% false positive rate and 31% false negative rate [812]; 74% accuracy [840].

PolyPhen2: 72% accuracy [578, 840]; 75% false positive rate when looking at ultra-conserved sites [841].

LRT: 28% false-negative rate and 6.7% false-positive rate [577].

MutationTaster: 85.7% accuracy [578]; 90.2% accuracy reported on the software’s website (<http://www.mutationtaster.org/info/statistics.html>).

GERP++: when plotting specificity against sensitivity, GERP++ had an area under the curve of 0.8921 for introns, 0.5976 for promoters and 0.8453 for coding regions [842].

It must be noted that measures of evaluation vary between studies, as they depend on which dataset the algorithm was compared to, the type of variants considered and the level of site conservation.

Variants predicted to be functional or non-functional by a single method are likely to have high false-positive rates and thus concordance of classifications from multiple algorithms is often considered more reliable and allows the user to harness the strengths of each method [588, 837, 843-845]. Databases such as ANNOVAR [846] and dbNSFP [847], which provide functional prediction scores for non-synonymous SNPs from various popular algorithms, facilitate comparison between methods. Even if concordance between multiple algorithms is relied on, however, it must not be forgotten that most algorithms focus on protein changes caused directly by a change in the DNA sequence, without considering the impact it may have on gene expression, for example through alternative splicing of mRNA [848, 849]. Furthermore, concordance between multiple algorithms may simply reflect the similarities in the methods upon which the predictions are made. For example, both SIFT and PolyPhen2 base their 'position-specific independent count' scores (a measure of the probability that a substituted amino acid will be tolerated) on multiple sequence alignments of the protein in question and related proteins; when scores differ, this may be dependent on technical details, such as which sequence alignments were used, rather than a fundamental difference in methods [850]. When relying on concordance between multiple algorithms, it is important to use algorithms that adopt a range of methods upon which to base their predictions. If all methods used the same criteria upon which to base their classifications, there would (or should) be no disparity between results and thus no

need to use more than one algorithm. As some scores are more correlated than others, one may have more confidence in a consensus between two scores that are less correlated than that between two highly correlated scores.

Consensus classifier algorithms have also been developed, which use scores from various pre-existing algorithms to obtain a consensus score, for example by taking a weighted average of the normalised scores of individual methods or by using a support vector machine approach. Although consensus classifier algorithms such as CONDEL, [851] CoVEC [852] and CAROL [840], have been shown to outperform individual methods when predicting the deleteriousness of variants, this is not always the case: for example, use of MutationTaster scores alone was shown to outperform CONDEL, in terms of true and false positive rates, when classifying pathogenic variants from the ExoVar dataset [853]. With the HumVar dataset, CONDEL outperformed all individual algorithms, but the area under the curve (0.88) was similar to that of PolyPhen2 (0.87) and MutationTaster (0.863). Even algorithms that use the same original scores to deduce a consensus score (albeit via different methods), such as CAROL and CONDEL, which both use original SIFT and Polyphen2 scores, can give conflicting results and may yield results that contradict the predictions from which they are derived [837].

There is no gold standard against which the performance of such algorithms, either consensus-based or concordance approaches, can be validated. There are many different databases of known disease-causing mutations, but they are often specific to a certain disease, such as the Amyotrophic Lateral Sclerosis Online Genetics Database (<http://alsod.iop.kcl.ac.uk/Index.aspx>) or Alzheimer Disease & Frontotemporal Dementia Mutation Database (<http://www.molgen.ua.ac.be/ADMutations>). Efforts have been made to collate findings from locus-specific databases, such as the Human Gene Mutation Database (<http://www.hgmd.org>), Human Genome Variation Database

(<http://gwas.biosciencedbc.jp>) and European Bioinformatics Institute Sequence Variation Database (<http://srs.ebi.ac.uk>). There is no consensus over which database should be used for validation; indeed, some may be more appropriate than others, depending on the phenotypic trait or variant in question. Validation results for a particular algorithm depend on the database to which it is compared [576]. It must also not be forgotten that such databases may not contain data on unpublished mutations.

6.1.4 Extreme phenotypes

Another way to increase power when investigating rare variant genotypic-phenotypic associations is to identify individuals who are most likely to have highly-penetrant rare mutations, for example, those with more severe disease types or more extreme responses to treatment. This method, known as selective genotyping or extreme trait phenotyping, is based on the assumption that rare variants have a stronger effect size than common variants (few common variants reported in the literature have an odds ratio of >2 [mostly between 1.1-1.4], whereas most rare variants have an odds ratio >2 , with a mean odds ratio of 3.74 [854]) and that the frequency of disease/trait-associated alleles are enriched in one or both groups with extreme phenotypes and so these individuals are genetically more informative [855, 856]. The extreme trait phenotyping approach has been used to identify an excess of rare variants in the *ABCA1* gene in individuals with low levels of high-density lipoprotein (HDL) cholesterol [857]. With quantitative traits, one can define extreme phenotypes by using cases whose phenotypes deviate from the population mean by a certain amount (those with unusually high or low trait values) [858]. Cases with trait values above or below the set thresholds can be treated as cases and controls [859], although this has been described as 'inefficient' as not all available data are used [858]. When using only cases with extreme phenotypes, a much larger source population size and recruitment cohort are needed to obtain the same numbers as used with standard case

control studies. However, given that sampling the extremes leads to higher power to detect associations, compared to random sampling, both for single-variant-based analyses [860, 861] burden analyses [861] and in family-based designs [862], this allows a smaller size to be used without compromising power. Accurate phenotyping is vital – as the sample size is reduced, even a small proportion of misclassified individuals could affect the analysis [863].

Considering the potential power gain when using extreme trait phenotyping, variant-pooling methods, and prioritising variants based on their likelihood of influencing disease risk or a phenotypic trait, predicted either by variant class or by prediction algorithms (the more variants are tested, the lower the p-value has to be to be declared statistically significant), it seems prudent to combine these methods. Theoretically, this would allow a causal variant/gene/pathway to be detected with an even lower number of cases.

Following on from the exploration of common variation and KD response with the candidate gene analysis and GWAS, an investigation of whether rare genetic variation across the genome influences KD response will be conducted. Various methods, including extreme trait phenotyping, variant prioritisation and gene- and pathway-based analyses, will be employed with the objective of maximising power.

6.2 Methods

6.2.1 Phenotypic data

The extreme trait phenotyping strategy was adopted in order to putatively identify those individuals most likely to have highly penetrant rare mutations that affect KD response.

Extreme responders and non-responders were included – a two-tailed selective genotyping approach. Two-tail selective genotyping has been shown to be more powerful than one-tail selective genotyping (when only one phenotypic extreme is selected for, and otherwise random sampling is used) [864], although a one-tailed approach is more appropriate when

the distribution of phenotypes is not normally-distributed, which suggests that cases in the dominant distribution arm are likely to be more genetically informative [865]. As shown in Figure 6.1, the distribution of responder status of the 108 (106 of which were genotyped with the Illumina chip) cases classified as extreme responders or non-responders (equivalent to the extreme summary response category, as outlined in Chapter 3), is not heavily skewed.

The criteria for classification as an extreme responder or non-responder is stringent (details are outlined in Chapter 3): extreme responders are those who achieved the most extreme reduction in seizure frequency and who maintained this seizure reduction over time and extreme non-responders are those who followed the diet for the shortest length of time with no beneficial effects, or with adverse effects on seizures; individuals included in this category are thus suitable for the extreme trait phenotyping strategy. Error in response classification is also reduced: if an individual reports a reduction in seizure frequency over a longer period of time (as opposed to a measure taken at just one time point) whilst following the KD, this is less likely to be due to chance.

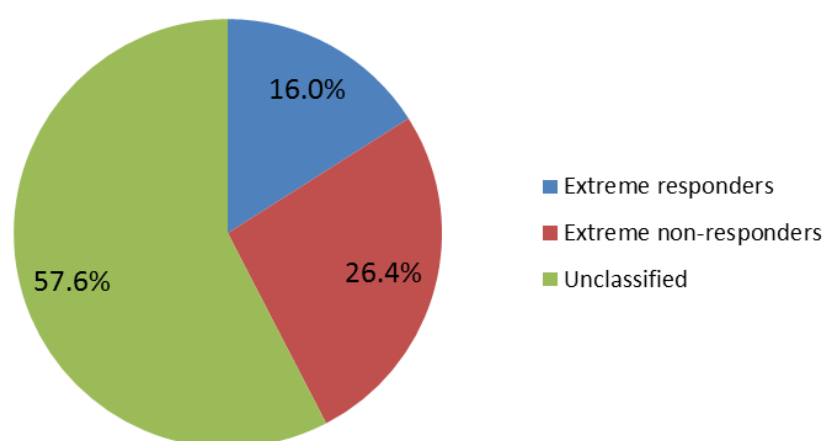


Figure 6.1: Extreme responders and non-responders, n=106 (out of 250 individuals with diet response data and genotyped with the Illumina chip)

6.2.2 Genetic data

Participants were genotyped with the *Infinium HumanOmniExpressExome Beadchip* (Illumina Inc, San Diego, USA). Such chips are a ‘halfway house’ between a standard genotyping array and exome sequencing and represent a powerful and cost-effective method of identifying rare variation in individuals and assessing association with phenotypic traits, as has been recently demonstrated [866].

6.2.3 Quality control and variant filtering

All markers present on the chip were subject to the same quality control procedures as employed in Chapter 5.

Per-individual quality control exclusion criteria:

Individuals with proportion of heterozygous SNPs <0.25 or >0.33 (command: *--het*).

Mismatch between gender reported in the phenotype file and gender imputed from PLINK (command: *--check-sex*).

Duplicate and related persons, by calculating IBD between all pairs of individuals (Pi-hat [proportion of IBD] >0.4 ; command: *--genome*).

Individuals with $>2\%$ missing SNP data (command: *--mind 0.02*).

Per-variant quality control exclusion criteria:

Cluster separation values <0.3 and Het-excess values between 0.1 and -1 and between 0.1 and 1 , given in GenomeStudio.

Genotyping rate <0.02 (command: *--geno 0.02*)

Deviation from HWE in the whole cohort: $p\text{-value} < 1 \times 10^{-6}$ (command: *--hwe-all --hwe 0.000001*)

All quality control-filtered variants were annotated using wANNOVAR (<http://wannovar.usc.edu>, [574]) and the hg19 reference genome. wANNOVAR is a web interface to the ANNOVAR software: a widely-used functional annotation tool that can help select variants most likely to be associated with disease or phenotypic traits. wANNOVAR has the advantage that it does not need ANNOVAR software or databases to be installed (which is very time-consuming), and it is easy to use.

The following information (representing all possible options available from wANNOVAR) was obtained for each variant:

- Variant class (exonic, intronic, intergenic, UTR and so on)
- Gene Name (using RefSeq gene definition)
- Exonic variant function (non-synonymous, synonymous and so on)
- Amino acid changes
- Region-level phastCons LOD scores
- Sequence identity score for the segmental duplication region in which the variant is located
- Alternative allele frequency in all subjects in the NHLBI-ESP project with 6500 exomes
- Alternative allele frequency data in 1000 Genomes Project (Feb 2012 release)
- dbSNP ID (taken from dbSNP135)
- Whole-exome SIFT scores for non-synonymous variants (based on Ensembl55 database)
- Whole-exome PhyloP scores*
- Whole-exome LJBsift (1-SIFT) scores*
- Whole-exome PolyPhen version 2 scores*
- Whole-exome LRT scores*
- Whole-exome MutationTaster scores*
- Whole-exome GERP++ scores*

* wANNOVAR/ANNOVAR obtains these scores from dbNSFP [847]

Based on the premise that rare variants are more likely than common variants to contribute to disease risk, variants were filtered by MAF according to 1000 Genomes Project Feb 2012 release and Exome Sequencing Project dataset (ESP6500). Variants with missing MAF information in ANNOVAR were excluded from analyses. Analyses were first conducted including all variants with $MAF < 5\%$. This threshold has been adopted in

publications to define uncommon or rare variants [867-869] and it complements the GWAS (see Chapter 5), which was designed to detect variants of $MAF > 5\%$. Further analyses were conducted with a more stringent definition of rare variation ($MAF < 0.5\%$), on the basis that the rarer the variant the more likely it is to be damaging. This definition of rare variation ($MAF < 0.5\%$) has been adopted in recent papers testing rare variant genotypic-phenotypic associations [797, 799, 866, 870-872].

Variants were not filtered according to MAF in this cohort. Variants, although rare in the general population (according to public databases), may not be rare in this cohort if one responder group is appropriately enriched for that variant, compared to the other group.

These variants (all rare variants, independent of variant class or prediction algorithm scores) will be referred to as group 0. These analyses account for the fact that synonymous SNPs or other non-coding variants may still affect gene behaviour – non-exonic splicing variants, for example, can lead to loss-of-function phenotypes [873]. When considered alone, these effects may be subtle but when considered alongside other variants, they may be critical [592]. It has been shown that filtering variants solely according to MAF or presence in public databases can be sufficient to detect a causal gene [874]. Furthermore, with SNP arrays, associations may be direct or indirect; it may be that a susceptibility variant (which may be synonymous, non-synonymous, coding or non-coding) is in LD with the susceptibility variant. Most common disease-associated variants found in GWAS are located in non-coding regions and are thought to affect regulatory elements; the same may be true for rare variants and so it is important not to discard them immediately from analyses.

Group 0 variants (both with $MAF < 5\%$ and $MAF < 0.5\%$) were further filtered to include only exonic (including exonic and exonic splicing) non-synonymous, stop-gain, stop-loss and

splicing variants. These variants are referred to as group 1. This is based on the premise that non-synonymous variants alter the protein coding sequence and are likely to be deleterious, thus increasing power to detect an association. If the associations were indirect, filtering variants in this way may not be of assistance – a non-synonymous variant may not necessarily be in high LD solely with other non-synonymous variants, but with variants of any class. However, the Illumina exome chip was designed to include functional exonic variants: it contains markers for variants associated with complex traits (such as diabetes, blood lipids and psychiatric traits) in previous GWAS, ancestry-informative markers (markers that showed strong differentiation between samples of African and European ancestry, and those of native American and European ancestry), markers for IBD estimation (markers that showed little differentiation between samples of African, European and Asian ancestry), a small set of ‘functionally interesting variants’, random synonymous variants, tags for human leukocyte antigen genes, coding variants in the mitochondria, a set of SNPs currently used as fingerprint SNPs at the University of Washington and Broad Institute, and a set of Y chromosome SNPs. With these variants, it is therefore more likely that a direct genotypic-phenotypic association is detected. Extracting coding variants that are more likely to be of functional consequence (non-synonymous variants) may increase power to detect an association and would facilitate interpretation of any association signals detected.

An additional analysis, solely with stop-gain and splice site variants, was considered a possibility. This analysis was not conducted as, akin to other non-synonymous variants, there is no guarantee that they result in loss-of-function or whether they are in LD with a variant that causes loss-of-function. Moreover, due to the small number of such variants present on the Illumina chip, power would be greatly reduced to detect a genotypic-phenotypic association.

Due to the uncertainty and variability associated with predictions from bioinformatics algorithms, which may lead to inclusion of neutral variants or exclusion of disease-associated variants, these will not be used to filter variants prior to analysis. They may, however, be used as a guide to estimate the effects of the most interesting variants (those that have the most differential distribution between responders and non-responders) on gene/protein function, although it should not be forgotten that variants on the Illumina chip (even if they were chosen to be on the exome chip, although this is less likely) may act as tag SNPs, rather than being the causal variants themselves.

6.2.4 Covariates and population structure

As outlined in Chapter 3, the effect of a range of demographic and clinical factors on KD response was assessed by logistic regression.

Association studies with rare variants are more susceptible to the effects of population stratification, leading to an increased risk of false-positive associations, compared to analyses with common variants. Rare variants are typically the result of recent mutations and so they are expected to be geographically clustered, or potentially only present in specific geographic populations (private alleles) [875]. This geographic clustering has been shown for variants with MAF <0.5%, both within European populations (Finnish populations had lower levels of allele sharing with other European populations than predicted by geographic distance) and, to a greater extent, across different continents [798]. Pairwise IBS distance was calculated between all individuals and MDS performed, as outlined in Chapter 5.

6.2.5 Association analyses

Binary files of quality control-filtered variants (group 0 and group 1), including phenotypes for all extreme responders and non-responders (excluding genetic outliers) were created in PLINK (v1.07, <http://pngu.mgh.harvard.edu/purcell/plink>, [558]).

Single variant association analyses were conducted in PLINK (command: *--fisher*), as outlined in Chapter 5. In all analyses, λ_{GC} was 1.00, indicative of absence of importance of population structure. Therefore, as in the common variant GWAS, MDS vector values were not included as covariates.

Analyses were first conducted including only quality control-filtered variants with MAF<5% (both with no filtering for variant class, and then just for non-synonymous and splicing variants) and then for variants with MAF<0.5% (with no filtering and then with filtering for variant class). P-values were obtained from 1000 permutations to adjust for multiple testing.

It has been shown that, even with a cohort consisting of approximately 7,000 cases and 17,000 controls, there is limited power to detect association signals using single variant tests for variants of MAF<0.5% with a modest effect size (odds ratio of 3) [797]. Because of the small sample size in this study, allele counts in responders and non-responders were examined, as well as p-values from the association tests. dbSNP (<http://www.ncbi.nlm.nih.gov/projects/SNP>) and Locuszoom (<http://csg.sph.umich.edu/locuszoom>, [696]), were used to determine whether the variants of interest were located within or were in LD with biologically plausible genes.

6.2.6 Gene- and pathway-based analyses

As in Chapter 5, variants were assigned to genes using wANNOVAR [574] and PLINK was used to conduct a gene-based set test and a self-contained pathway-based test. A Bonferroni-adjusted significance threshold was set by dividing 0.05 by the number of genes or pathways included in the analysis.

Variants with MAF<5% with no filtering for variant class were used as, although synonymous variants or variants of unknown function are often hard to interpret when

examined individually, they may play important roles in specific networks or pathways; using the less stringent MAF filtering also avoids exclusion of potentially interesting variants/genes.

6.3 Results

6.3.1 Quality control and variant filtering

In total, there were 950976 makers on the Illumina chip, 81965 of which had a MAF<5% in 1000 Genomes Project Feb 2012 release or Exome Sequencing Project dataset (ESP6500). 48117 variants had a MAF<0.5%. The function and, where applicable, exonic function of these variants (information taken from wANNOVAR output file) is given in Table 6.1.

Table 6.1: Function of variants included in the Group 0 MAF<5% and MAF<0.5% analyses (prior to per-SNP quality control filtering)

Variant function	Exonic function	Number of variants present on chip minor allele frequency<5%	Number of variants present on chip minor allele frequency<0.5%
Downstream	n/a	360	36
Exonic	Non-synonymous	53847*	43919*
	Stop-gain	213*	197*
	Stop-loss	149*	131*
	Synonymous	1976	1326
	Unknown	379	282
Exonic splicing	Non-synonymous	475*	432*
	Stop-gain	2*	2*
	Synonymous	248	202
	Unknown	3	2
Intergenic	n/a	11716	317
Intronic	n/a	10062	509
Exonic non-coding RNA	n/a	81	19
Intronic non-coding RNA	n/a	591	4
Splicing non-coding RNA	n/a	0	0
UTR3 non-coding RNA	n/a	16	2
UTR5 non-coding RNA	n/a	1	0
Splicing	n/a	591*	535*
Upstream	n/a	302	26
Upstream; downstream	n/a	15	3
UTR3	n/a	836	148
UTR5	n/a	102	25
Total		81965	48117

*variants included in Group 1 analyses (missense, nonsense and splicing variants)

After per-SNP quality-control filtering, 79169 SNPs remained in Group 0 MAF<5%. 31 markers failed the HWE test ($p\text{-value}<1\times10^{-6}$). 2773 SNPs failed the missingness test ($\text{geno} > 0.02$).

After per-SNP quality-control filtering, 46517 SNPs remained in Group 0 MAF<0.5%. 4 markers failed the HWE test ($p\text{-value}<1\times10^{-6}$). 1597 SNPs failed the missingness test ($\text{geno} > 0.02$).

After per-SNP quality-control filtering, 53409 SNPs remained in Group 1 MAF<5%. 12 markers failed the HWE test ($p\text{-value}<1\times10^{-6}$). 1861 SNPs failed the missingness test ($\text{geno} > 0.02$).

After per-SNP quality-control filtering, 43720 SNPs remained in Group 1 MAF<0.5%. 4 markers failed the HWE test ($p\text{-value}<1\times10^{-6}$) and 1493 SNPs failed the missingness test ($\text{geno} > 0.02$).

Following quality-control filtering, 66 non-responders and 38 responders remained in association analyses.

There was no strong evidence to suggest that KD response was affected by any clinical or demographic factors (see Chapter 3) and so none were included as covariates.

6.3.2 Association analyses

6.3.2.1 Single variant approach

Group 0:

No statistically significant results were obtained from the single variant association test MAF<5%. rs12565493 and rs12192673 had the lowest p-value (unadjusted p-value=0.0019, p-value obtained from 1000 permutations=0.96).

rs7969998 had the largest differences in minor allele count between responders and non-responders, with a frequency of 0.1705 in non-responders and 0.01923 in responders.

The variants with the highest case- or control-unique minor allele counts were rs12565493 and rs12192673, with a frequency of 0 in non-responders and 0.1154 in responders, and rs752318, rs10038594, rs2039979, rs11788407, rs17593299 and rs11076243, with a frequency of 0.1222 in non-responders and 0 in responders.

rs12565493 is located in the intronic region of *FAF1*.

rs12192673 is located in the 3'UTR of *GNL1*.

rs7969998 is not located within any gene. It is in an LD region with *THAP2*, *TMEM19*, *RAB21*, *TBC1D15*, *MRS2P2*, *TPH2* and *ZFC3H1*.

rs17395014 is located in the intronic region of *CACHD1*.

rs752318 is located in the intronic region of *PLXDC2*.

rs10038594 is in an LD block with *PLK2*, *GAPT* and *RAB3C*.

rs2039979 and rs11788407 are in an LD block with *ANXA1*.

rs17593299 is not in an LD block with any gene.

rs11076243 is located in the intronic region of *NDRG4*.

Many of these genes play a role that could feasibly influence KD response. For example, interaction of *FAF1* with the FAS antigen (TNFRSF6) mediates apoptosis. *THAP2* (THAP domain containing, apoptosis associated protein 2), by its name, is also associated with apoptosis.

RAB21 encodes a GTPase that plays a role in trafficking of integrins and so regulates cell adhesion and migration. *TBC1D15* encodes a brain GTPase activating protein that is thought to regulate intracellular trafficking. Both *GNL1* and *RAB3C* have the GO term 'GTP binding'. Variants in *RAB3C* and variants located between *RAB3C* and *GAPT* have been associated with insulin resistance [713].

TPH2 is predominantly expressed in the brain stem and encodes a protein that catalyses the first step in serotonin synthesis. Mutations in this gene have been associated with autism [876], attention-deficit/hyperactivity disorder [877], obsessive compulsive disorder [878], major depression [879, 880] and susceptibility to suicide [881, 882].

CACHD1 encodes a protein that is thought to regulate voltage-dependent calcium channels.

Annexin I, encoded by *ANXA*, may have antiinflammatory activity as it inhibits phospholipase A2.

NDRG4 encodes a protein that contributes to the homeostasis of intracerebral BDNF levels.

The MAF<0.5% analysis generated higher p-values than the MAF<5% analysis, and the variant with the lowest p-value in the MAF<0.5% analysis had exactly the same p-value in the MAF<5% analysis. This indicates that no power was gained from excluding variants with MAF 0.5-5%.

exm237818, exm544941 and exm648484 had the lowest p-value (unadjusted p-value=0.047; p-value obtained from 1000 permutations=0.95) and the highest case- or control-unique minor allele counts, with a frequency of 0 in non-responders and 0.058 in responders. These variants are located in *SCN2A*, *DAAM2* and *ATXN7L1* respectively.

The variants with the greatest difference in minor allele counts between responders and non-responders were rs923296 and exm1581807, both with a frequency of 0.044 in non-

responders and 0.019 in responders. rs923296 is located in an LD block with many genes (*HHIPL2*, *TAF1A*, *MIA3*, *AIDA*, *BROX*, *FAM177B*, *DISP1*, *TLR5*, *SUSD4* and *C1orf65*) and it is difficult to know which gene is relevant; exm1581807 is located in *PCNT*.

SCN2A (sodium channel, voltage-gated, type II, alpha subunit) may be biologically relevant with regards to influencing KD response. Mutations in *SCN2A*, a sodium channel gene which is expressed in the brain, have been associated with several seizure disorders, including benign familial neonatal infantile seizures, febrile seizures plus, and intractable epilepsy of infancy [883-885]. rs2228980 was found in three responders, all with the genotype A/G. Two of the responders were thought to have epilepsy with myoclonic atonic seizures (one at the 'severe end of the spectrum' and one with 'probable myoclonic astatic epilepsy') and one had an unspecified epilepsy syndrome (according to clinic letters). This is a missense variant, which has a GERP++ score of 5.14 and is classified as damaging by LRT, but is classified as tolerated by SIFT, benign by PolyPhen2 and as a polymorphism by MutationTaster. rs2228980 may be relevant to the cause of epilepsy in these individuals and this type of epilepsy may respond well to the KD.

DAAM2 has the GO term 'Rho GTPase binding' and *ATXN7L1* has been associated with Alzheimer's disease [886].

Group 1:

When including all non-synonymous or splicing variants with MAF<5% in the single variant association test, no statistically significant results were obtained. exm255464 and exm799515 had the lowest p-value (unadjusted p-value=0.0058; p-value obtained from 1000 permutations=0.91).

Aside from these variants, the variants with the largest difference in minor allele counts between responders and non-responders were exm887741 and exm1204532, which had a frequency of 0.11 in non-responders and 0.019 in responders.

The variants with the highest case- or control-unique minor allele counts were exm1188563, with a frequency of 0 in responders and 0.11 in non-responders, and exm255464 and exm799515, both with a frequency of 0 in non-responders and 0.009 in responders.

exm255464 is located in *RFTN2*.

exm799515 is located in *TMEM141*.

exm887741 is located in *OLFML1*.

exm1204532 is located in *GFER*.

exm1188563 is located in *WDR93*.

None of these genes have an obviously relevant biological function with regards KD response.

In the Group 1 MAF<0.5% analysis, exm237818, exm544941, exm648484 had the lowest p-value (unadjusted p-value=0.04731; p-value obtained from 1000 permutations=0.93), the same variants with the lowest p-values in the Group 0 analysis. Variants with the largest difference in minor allele count between responders and non-responders (rs923296 and exm1581807) and the highest case- or control-unique minor allele counts (exm237818, exm544941, exm648484) were noted in Group 0 analyses, as above.

6.3.2.2 Gene and pathway-based approaches

A significance threshold of 3.21×10^{-6} (0.05/15589) and a suggestive significance threshold of 6.41×10^{-5} (1/15589) were set for the gene-based tests. A significance threshold of

3.79×10^{-5} (0.05/1320) and a suggestive significance threshold of 0.00076 (1/1320) were set for the pathway-based tests.

No gene or pathway reached significance or suggestive significance. The gene with the lowest p-value ($p=0.0019$) was *MCMBP* and the pathway with the lowest p-value ($p=0.0019$) was the PID 'S1P4' (Sphingosine 1-phosphate type 4) pathway.

The following GO Biological Processes are associated with *MCMBP*: DNA-dependent DNA replication, mitosis, mitosis S phase and sister chromatid cohesion. The following GO Molecular Processes are associated with *MCMBP*: chromatic binding and protein binding.

The S1P4 pathway contains *CDC42*, *GNA12*, *GNA13*, *GNAI1*, *GNAI2*, *GNAI3*, *GNAO1*, *GNAZ*, *MAPK1*, *MAPK3*, *PLCG1*, *RHOA*, *S1PR4* and *S1PR5*. This pathway is shown in Figure 6.2.

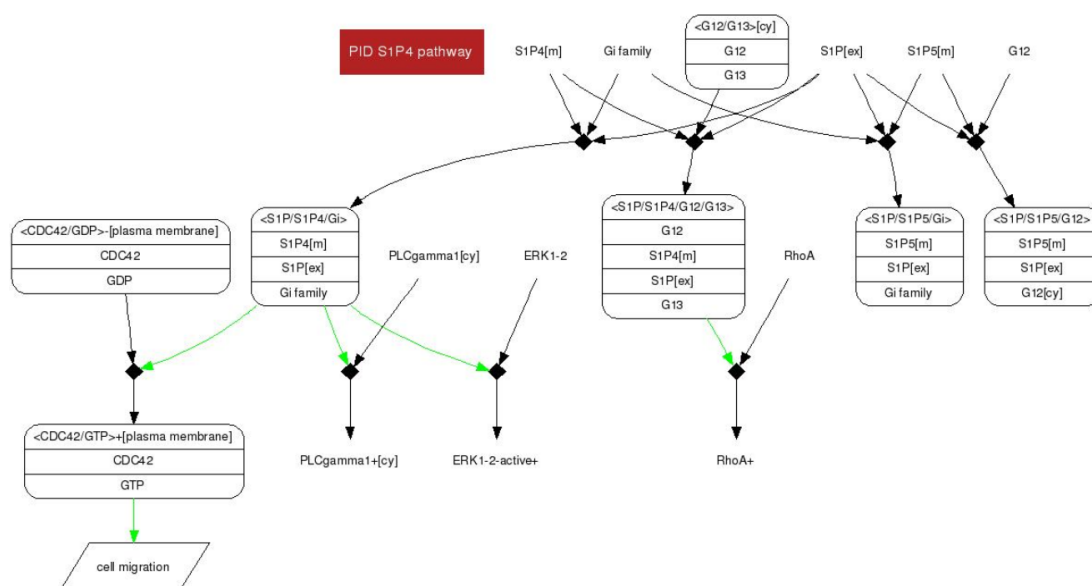


Figure 6.2: The Pathway Interaction Database S1P4 pathway, taken from <http://pid.nci.nih.gov>

6.4 Discussion

Key findings:

- No significant results were obtained from single variant, gene-based or pathway-based association analyses, although some common themes pervade that could feasibly account for variability in KD response.
- Filtering for rare, non-synonymous and splicing variants did not improve power in association tests.
- The lack of significant results may be due to a number of reasons: lack of power to detect an association, using an array to detect rare variants, or simply because rare variation does not influence response to the KD.
- Certain genes with variant(s) that have the greatest difference in minor allele count in responders and non-responders could feasibly influence KD response.

Interpretation is difficult as only a few individuals harbour variants in these genes.

6.4.1 Power: study design and variant prioritisation

The principal reason why this study was underpowered to detect a rare variant genotypic-phenotypic association was due to the sample size, which was even further reduced by only including participants with an extreme response to the KD. The rationale for this was to increase power by selecting those individuals thought to be most genetically enriched for variants associated with the phenotype.

Use of a quantitative, rather than a dichotomous, phenotypic trait would have increased power and enabled determination of a quantitative relationship between any gene/variant found to be associated with the phenotype [887]. The difficulties of achieving this when clinic letters are the predominant data source, has been discussed in previous chapters.

The aim of narrowing down the number of variants to exclude those that were more likely to be phenotypically neutral was to increase power. Most analyses focus on protein-altering variants, as these are the most amenable to functional interpretation. As outlined in the Introduction of this chapter, when prioritising non-synonymous variants (or loss-of-

function or whatever criteria are chosen), it is not guaranteed that one ends up with a functionally-enriched subset of the genome and one risks losing any causal synonymous or non-coding variants. In this study, group 1 analyses generated higher p-values than group 0 analyses, which is most likely to be due to the reduced number of variants. Using bioinformatics tools to filter variants is also fraught with uncertainty: they may not always reflect up-to-date versions of their source resources, such as public SNP databases [850] and, at least when using software that provide multiple sources of information, such as wANNOVAR and dbNSFP, no reliability measures are provided for each prediction. There is a need to continue developing methods to robustly identify functionally important variation.

Due to the ambiguity that filtering variants entails, in this study, analyses were also conducted with variants purely filtered based on MAF (Group 0). Two definitions of rare variation were adopted, one less stringent than the other. A definition of $MAF < 0.5\%$ as a hard filter could be considered too stringent to adopt and it may have excluded some potential variants of interest. A risk allele that affects a certain trait in its homozygous form may exist at a relatively high frequency in a heterozygous form in the general population. Also, there may not be a strong negative selection pressure for variants that influence KD response, as they may not affect fitness in the general population. For example, survival in the face of starvation may not be relevant for certain populations and variants that adversely affect KD response may be important for another phenotypic trait. Finally, a relaxation in natural selection in recent generations may favour the accumulation of deleterious rare alleles in the population [888, 889]. Analyses were also conducted excluding variants with $MAF > 0.5\%$ as, if the common disease/rare variant hypothesis is true in the case of KD response, it could be assumed that very rare variants are more likely to have a large phenotypic effect. Several variants examined from the $MAF < 0.5\%$ analyses

were of interest, and variants such as those in *SCN2A*, which are highly relevant to this population, would have been missed if only the top results from the MAF<5% analyses were examined (no significant results were obtained in any analysis and so only those variants with the lowest p-values or the greatest difference between responder/non-responder minor allele counts were examined). It is a vicious cycle: the less filtering used in the first place, the higher the multiple testing burden, which reduces the likelihood of obtaining statistically significant results. No significant results were obtained in any analysis (each with a different grade of variant filtering) and so it seems unlikely that 'suboptimal' filtering was the cause of the lack of significant findings.

There are various alternative ways either to prioritise variants prior to analysis or to aid interpretation of results. One example of this is consideration of gene expression profiles in tissues thought to be relevant to the phenotype. The most likely candidates for this with regards KD response would be the brain, liver and pancreas, considering the role of these organs in fatty acid and glucose metabolism and seizure onset. This is not to say that genes expressed elsewhere may not be of importance. For example, the kidneys, due to their role in glucose and ketone excretion, may influence KD response, as may ubiquitously-expressed fatty acid/carnitine/glucose transporters. Indeed, the main known genetic influence on KD response, *SLC2A1*, is expressed in the endothelial cells of normal brain, but also ubiquitously expressed in most cells; it is the most widely expressed of the hexose transporters [890]. Based purely on expression levels, one may expect *SLC2A3* (encoding GLUT3) to be more relevant to KD response, due to its specific expression in neurons [891]. Considering this is the first genetic analysis to be conducted with regards KD response, it seems prudent to use such strategies to aid interpretation of results, rather than to exclude genes/variants in the first place.

Another option would be to prioritise candidate mutations by quantifying gene intolerance to functional mutation, using a residual variation intolerance score [892], which can be used by itself or in combination with a variant-level prediction of functionality, such as those adopted in this chapter (GERP++, SIFT, PolyPhen2 and so on). This score is based on the premise that genes intolerant to mutation are less likely to harbour damaging mutations, unless they are contributing to disease, which was shown to be true for genes that cause Mendelian disorders – they are more likely to have a lower residual variation intolerance score, reflective of lower tolerance to mutation. However, whether genes are more likely to influence a phenotype if they do not tend to harbour functional variants in the general population or whether they commonly harbour functional variation, seemingly with no clinical association, depends on the phenotypic trait in question. This was shown, for example, with genes linked to developmental disorders, which were more likely to be caused by genes that were intolerant to functional variation, and immunological disorders, which were more likely to be caused by genes with an excess of common functional variation, compared to the genome-wide average number of common functional variants found in genes with a similar burden of mutations. It is unknown whether it would be best to prioritise genes with low or high intolerance scores and so it would not make sense to filter variants/genes prior to analysis, based on this strategy. Furthermore, one cannot prioritise variants using a large number of strategies until the desired results are obtained, as this increases the multiple testing burden and could be considered as manipulation of results. Until a gold standard is set for prioritising or filtering variants/genes, one must decide on a particular strategy, or at least limit the number of strategies adopted. By virtue of the lack of statistically significant results in any analysis, none of the aforementioned strategies were adopted to interpret results.

The fact that λ_{GC} was equal to 1.00 in all single variant analyses indicates that results were not biased by population stratification.

6.4.2 Results of interest

Independent of statistical significance, some interesting results were obtained when looking at single variant, gene-based and pathway-based analyses. Some variants with the greatest difference in responder/non-responder minor allele counts or with case- or control-unique variants were located in genes that could feasibly be linked with KD response. Several genes play a role in apoptosis, a process which has been proposed as a mechanism of action of the KD. *FAF1*, for example, is a member of the cell death-inducing signalling complex mediated by FAS (it binds directly to FAS and caspase-8) [893], and it suppresses NF- κ B activity, which has a downstream effect on expression levels of genes involved in the immune response, inflammatory processes and apoptosis [894]. It has been postulated that *FAF1* is related to the pathogenesis of neurodegenerative diseases: gene expression was reported to be increased in the brains of people with Parkinson's disease and Alzheimer's disease [895] and *FAF1* deficiency blocked biochemical events that are linked with these disorders, including caspase and JNK activation, ROS generation and cell death [896].

Other genes of interest in these analyses are involved in neurological processes, such as serotonin synthesis and regulation of BDNF levels. Both may be relevant to KD response: serotonin may influence seizure susceptibility [304] and concentrations were shown to be reduced in the CSF of children following the KD, compared to pre-diet levels, although concentrations did not differ between responders and non-responders [378]; as BDNF-mediated activation of TrkB pathways have been shown to promote hyperexcitability and kindling [897, 898], if reduced glucose levels from KD-feeding inhibit expression of BDNF (this has been shown in animal models where glycolysis was interrupted with

administration of 2DG [237]), this would be a plausible mechanism for reducing seizures [233].

Another predominant theme was genes encoding GTPases, or GTPase-activating or -binding proteins. Some genes, such as *TBC1D15*, may be relevant to KD response: TBC1D15 is a Rab7-selective GTPase-activating protein (it inactivates Rab7) and its expression leads to lysosome fragmentation and decreases cellular sensitivity to growth factor withdrawal-induced apoptosis [899, 900]. Dysregulation of specific Rab (including Rab7) has been postulated as a marker for neuronal dysfunction, due to the association of upregulation of these genes with Alzheimer's disease mild cognitive impairment.

Variation in *SCN2A* may be relevant to the cause of epilepsy in certain individuals. A missense transcript variant (K908R) at chr2:166201225 was found in three responders (in addition to many other synonymous or non-coding variants, which were present in individuals in both response groups). No publications were identified that reported an association between variation at this locus and seizures, but it has been patented for a genetic assay for determining predisposition to epilepsy, following discovery of the mutation in an individual with idiopathic generalised epilepsy [901] (patent publication number US7485449 B2). The mutation is thought to be important as it is located in an extracellular domain between two transmembrane domains that form the wall of the transmembrane pore, which allows the sodium to enter the cell. The phenotypes of two out of the three individuals in which this variant was found fit into the spectrum of epilepsy with myoclonic atonic seizures or generalized epilepsy with febrile seizures plus, which have been associated with *SCN2A* mutations [902-904]. The other individual was described as having an 'unspecified epilepsy syndrome' with two types of seizures: one (the predominant seizure type) where the patient becomes stiff and suddenly jerks all over, lasting for 2-3 minutes, and another where the left arm raises and the eyes deviate to left,

which occur in clusters. These results highlight the potential to unearth genotypic relationships with phenotypes other than that which is principally being investigated, such as cause of epilepsy. They may not be clinically useful when predicting likelihood of KD response in patients but, if the mutation is confirmed with sequencing and found to be of functional importance, it may still be of interest to patients and their families with regards the cause of their epilepsy, regardless of whether it changes clinical management or not. It is possible that more discoveries of this nature could have been made if personalised markers (for example, variants associated with known epilepsy-candidate genes) had been added to the chip.

MCMBP, the gene with the lowest p-value in the gene-based analysis, encodes a protein that regulates initiation of DNA replication and elongation. For DNA to replicate, so-called 'pre-replicative complexes' are assembled during the G1 (Gap 1 – the first stage of interphase) phase of the cell cycle; these pre-replicative complexes are activated at the S phase of the cell cycle [905]. The minichromosome maintenance (MCM) complex is an essential component of pre-replicative complexes. The MCMBP protein can bind to any of the subunits within the MCM complex, interacting most strongly with MCM4 and MCM7 [905]. Phosphorylation of MCM4 by cyclin A/Cdk2 regulates the activity of the MCM complex [906], as does protein interaction with MCM7 [907]. Deletion of the *MCMBP* ortholog gene in yeast leads to cell cycle arrest and overexpression of the gene inhibits DNA replication and causes DNA damage [908]. MCM proteins have also been found to regulate hypoxia-inducible factor 1 (HIF1) activity (a gene that mediates changes in gene expression in response to hypoxia, including [among many others] regulation of GLUT1): MCM proteins were shown to directly interact with HIF1 α and negatively regulate HIF1 activity, which prevents the induction of cell cycle arrest that is normally induced by HIF1 when oxygen is in low supply [909]. When oxygen is not in low supply, MCM proteins

promote cell cycle progression by inhibiting HIF1 but, in situations of hypoxia, MCM expression is downregulated and the activity of HIF1 works to inhibit cell cycle progression (otherwise, this would exacerbate the difference between oxygen supply and demand). It is feasible that variation in *MCMBP* affects cell cycle progression and other functions of the MCM complex, such as the response to hypoxia. *MCMBP* may influence KD response via regulation of cell cycle progression and cell survival/death in neurons.

The most closely associated pathway with KD response was the S1P4 pathway. S1P4, one of the S1P (Sphingosine-1-phosphate) receptors, is predominantly expressed in cells and tissues of the lymphoid system and is thought to play a role in immune response [910]. It has been shown to activate Gα12/13 and the small GTPase Rho; overexpression or ectopic expression of *S1P4* lead to increased cell motility and rearrangements in the cytoskeleton [911]. Binding of S1P4 to S1P activates the MAPK ERK [912] and phospholipase C (PLC) downstream pathways [913, 914], both of which are involved in many processes, including cell proliferation [915, 916], apoptosis [917, 918] synaptic plasticity/the BDNF cascade [919, 920], and insulin sensitivity/glucose transport [921-924]. Binding of S1P4 to S1P has also been found to induce cell migration via activation of CDC42 [925], a member of the Rho family of small GTPases that regulates cell cycle progression, induces apoptosis in response to stress in neurons [926, 927] and is essential for insulin secretion [928, 929].

As outlined in Chapter 5, the MAPK and other pathways related to cell cycling could potentially play a role in seizure control due to effects on neuronal survival/death. The effects on glucose-stimulated insulin secretion represent an alternative way through which the KD may exert its antiseizure effects, although the two mechanisms of action are not necessarily mutually exclusive.

These observations suggest that certain genes may influence KD response in particular individuals. It may be the case, for example, that individuals with certain types of epilepsy (with specific genetic causes) respond in a particular way to the KD. Variation in the S1P4 pathway may impair or enhance the effects induced by the KD that lead to seizure reduction. Depending on how many individuals the variant(s) were present in, this would not necessarily be picked up with an association analysis, nor when examining those variants or genes with the largest difference in responder/non-responder minor allele counts. For example, one non-responder may harbour a variant that adversely affected his/her response to the KD but, as it was only present in one individual, it would not generate a low p-value in association analyses; examination of all genes in which one responder or non-responder harbour a variant would not be feasible and there would be no way of determining whether these variants/genes actually influenced KD response or not. Genes with related functions may overcome, or at least partially overcome, the impaired function of certain genes. Larger sample sizes are needed in order to identify more individuals with rare variants, but interpretation will still be complicated by the many complex mechanisms underlying the antiseizure effects of the KD.

6.4.3 Rare variant detection with arrays versus sequencing

The Illumina chip used for these analyses, which is a combination of a SNP array and an exome chip, allows a higher frequency of rare, putatively functional variants across the genome to be typed, compared to standard SNP arrays, which typically provide information (through LD correlation) for 80-90% of common variation (MAF>5%), far less for variants of intermediate frequency (MAF 0.01-5%) and virtually none for rare variants [500]. Such chips still rely on LD mapping – any association signal found may be the true causal variant or may just be in LD with it. With rare variants, indirect association mapping is not ideal, as LD tends to be weak between common tag SNPs and rare variants [862, 930].

Considering that there are more low MAF variants typed on this Illumina chip compared to standard SNP arrays, it is more likely that a higher number of rare variants (with similar MAFs to the typed variants) will be highly correlated, but even rare variants may not be in strong LD with rare SNPs [799]. Although putative functional exonic variants were included on the chip, there is no guarantee that these variants affect KD response, whether they lie within genes that affect KD response, or whether they are in LD with variants that affect KD response.

Sequencing is the preferred method for detection of rare variants, as it only relies on direct association mapping and coverage is obviously more complete, allowing for analysis of very rare or private variants. Extremely rare variants will either not have been seen before or have such low allele frequencies that they have not been included in commercially available genotyping chips [714]; with the exome chip, non-synonymous variants had to be observed at least three times in two or more studies, and splicing and stop-altering variants had to be observed at least twice in two or more studies

(http://genome.sph.umich.edu/wiki/Exome_Chip_Design). A further advantage of sequencing is that it provides more information on structural genetic variation, as opposed to the exome chip, which contains only 181 indels and is unable to address frameshift mutations.

Despite these drawbacks of exome array genotyping, genotypic-phenotypic associations have been reported using this technology. In a study including over eight thousand participants, two variants in genes at known loci and three new genes were shown to be associated with either fasting proinsulin concentrations or insulinogenic index (the ratio of the increase in insulin concentration to the increase in plasma glucose after a glucose load) [866]. In another study, including almost 3,000 breast cancer cases and over 7,000 controls, rs145889899 showed a statistically significant association with breast cancer and was

almost exclusively seen in African Americans. No statistically significant results were obtained for prostate cancer [714]. These studies serve as a proof-of-concept for rare variant genotypic-phenotypic detection using array-based technology, but they also highlight the need for large sample sizes. This is easier to achieve with arrays, compared to whole exome or genome sequencing, due to lower prices.

6.4.4 Rare variant/common disease hypothesis

Considering the results from these analyses, it is difficult to conclude whether rare genetic variation influences KD response, due to the small cohort size. In order for rare variants to play an important role in explaining the missing genetic contribution to disease/a phenotypic trait and be detectable in association studies, they must have larger effect sizes than those observed for common variants. If one concentrates on coding variants, the effect sizes have to be even higher, as variants are selected from a very small proportion of the genome [714]. It may be that genetic variation (whether it be common or rare) contributes to KD response but the effect sizes of each variant are too small for associations to be detected. This is particularly likely to be the case in studies such as this, with a small cohort. Effect sizes are likely to be higher when variants are grouped together, for example, by GO term or pathway, but results from this approach were hard to interpret in this study.

Structural variants, such as indels, larger CNVs, translocations and inversions, which are incompletely captured by genotyping arrays, as well as undetected gene-gene or gene-environment interactions may also play a role in the genetic contribution to disease (for example, the expression of one gene may be altered by another loci or by the environment, which may results in a different effect on the phenotype) [931, 932], or, in this case, KD response. Sequencing can help with the matter of structural variation, and aggregate analyses that consider groups of genes in a pathway or gene-gene interactions may help

elucidate the complex and as-of-yet undiscovered genetic contribution to KD response. Incorporating the effects of environmental interactions into genetic models is problematic, principally because it is unknown what exactly the relevant environments are. One may stratify association studies by environment, but this would likely only reveal the largest interaction effects [932].

6.4.5 Conclusions and future work

Using single variant, gene-based and pathway-based association tests, no significant associations of rare genetic variation with KD response were found. When examining responder/non-responder minor allele counts, some biologically plausible genes were identified that may at least partly explain KD response or even the cause of epilepsy in certain individuals. Some genes of interest share common functions; taken together with results from the pathway analysis, it seems that rare variation affecting cell cycle progression, cell survival/apoptosis, neurological processes and insulin sensitivity contribute to KD response. Due to the fact that no single gene can be pinpointed as the reason for extreme KD response (whether favourable or unfavourable) in this cohort, this makes clinical translation of results difficult.

The lack of significant results may be explained by a number of reasons, including small sample size, small effect size of rare variants, the number of variants in a gene and their frequency, and the indirect association mapping approach of the genotyping array; even though the Illumina chip contains a higher number of rare, putative functional variants than standard SNP arrays, discovery of any potential genotypic-phenotypic associations is limited by the content of the array.

If there is no prior knowledge regarding the type of variant (for example, whether it is present in a conserved region or not) or properties of genes that are likely to affect the phenotypic trait in question, as is the case with KD response, it may be best to adopt less

stringent filters, or indeed no filters at all, and then use information such as variant class and predicted deleteriousness to aid interpretation of results post-analysis. Even with the knowledge that functional variants are more skewed towards the lower end of the MAF spectrum [933], it is unknown whether variants that affect KD response also tend to be rare, or whether they are subject to purifying selection at all (especially when we are dealing with response to treatment, rather than a disease state *per se*).

Ideally, sequencing, where the matter of incomplete LD between tag and causal variants is not an issue, would be used to investigate the effects of rare variants on KD response. This is the focus of the next chapter.

7 Exonic variation and response to the Ketogenic diet: whole exome sequencing

7.1 Introduction

Unbiased sequencing approaches – whole genome or whole exome sequencing (targeted sequencing of all coding regions of the genome) – have become increasingly available, both for family-based studies and for groups of unrelated individuals, since the commercial release of a next-generation sequencing instrument in 2005, at increased speeds and reduced costs compared to traditional Sanger sequencing [934, 935]. Technologies such as Illumina Genome Analyzer or HiSeq, Life Technologies SOLiD, or Roche 454 Genome Sequencer [936], generate sequence data (generally short reads of 35-450 base pairs [937]) from hundreds of millions of spatially segregated, amplified DNA templates in parallel [938, 939].

The often prohibitive cost of whole genome sequencing means that use of whole exome sequencing has thus far been more widespread. With exome sequencing, the DNA sample is reduced or enriched, leading to a sample where the content is made up of the protein coding regions of the genome, along with adjacent intronic content such as UTRs and miRNA binding sites. This is most commonly achieved by PCR, which takes a single-stranded piece of DNA as a primer, obtained by denaturing the DNA with heat, and amplifies specific DNA sequences through use of enzymes and probes attached to targeted regions, using the original strands as templates. Capture will never be 100% complete, as some regions of interest resist the enrichment strategy. This is followed by high-throughput next-generation sequencing of the exon-enriched sample.

Currently, most exome sequencing success stories have been for Mendelian disorders, such as congenital chloride diarrhoea (*SLC26A3*) [940], Miller syndrome (*DHODH*) [941] and non-syndromic hearing loss (*GPSM2*) [942]. Such mutation identification is important for those who desire a diagnosis, even if it does not necessarily lead to a change in disease treatment and management. There is now hope that exome sequencing will be of use in unearthing the genetic basis of complex traits, for which genetic variation may *influence* disease risk or a phenotypic trait, rather than *cause* it [943-945]. Most commonly, individuals at the extreme ends of the phenotypic spectrum are studied; this approach will most probably remain common until it is routine to sequence many thousands of samples in one study. The first example of this approach was in 2010, when sequencing the exomes of patients with combined hypolipidaemia (individuals with very low triglyceride, LDL and HDL cholesterol levels) identified two nonsense mutations in *ANGPTL3* that, when both nonsense alleles are present, lead to combined hypolipidaemia [946].

Exome sequencing allows investigation of both common and rare coding variation, with a relative excess of rare variation [944]. This rare variation can partly be attributed to recent population expansion [616, 947, 948] and also because rare variants in protein-coding regions are under purifying selection and are thus more likely to be deleterious [948, 949]. In comparison to genotyping arrays, where rare variant investigation is made problematic by the fact that, due to their low MAF, rare causal variants are less likely to be in high LD with (and therefore tagged by) nearby markers, the direct association mapping provided by sequencing allows discovery of rare, potentially private variation and the opportunity to explore their phenotypic effects. Not limited just to single point mutations, sequencing also permits discovery and exploration of indels and larger structural variants, such as copy number variants (CNVs - defined as DNA segments of ≥ 1 kb in size present at variable copy number in comparison with a reference genome [950]), which are known to be prevalent

and can be of functional consequence in humans [951]. Indels are the most abundant form of genetic variation in humans after SNPs [825] and both frameshift and in-frame indels have been linked with various diseases, such as cystic fibrosis [952]. Structural variants affect a higher proportion of the genome than do SNPs [953, 954] and they have been associated with various diseases, such as autism [955, 956], schizophrenia [957, 958] and Crohn's disease [827, 959].

Considering that many Mendelian genetic disorders are caused by mutations in protein-coding sequences [874, 961], whole exome sequencing is a powerful approach for exploring rare genotypic-phenotypic associations as, theoretically, the sample is enriched for variants of functional consequence. Due to its lower cost compared to whole genome sequencing, more samples can be sequenced at higher depths, leading to higher accuracy in rare variant calls [962]. Potentially important regulatory variation present in non-coding regions will, however, be missed [963]. This may be of consequence for complex traits, where non-coding genetic variation is believed to play a larger role than in Mendelian disorders [944]. Characterisation of such non-coding variation may be difficult as their effects are often unknown, making it harder to pinpoint causative genes [964].

Furthermore, considering that the effect sizes of non-coding variants are thought to be smaller than those of coding variants, studies may be underpowered to detect phenotypic associations [944].

7.1.1 Accurate variant detection

One of the main challenges with next-generation sequencing data is accurate variant detection. A mismatch when aligning the sequence reads against the reference genome could represent a true SNV, but could also be the result of an error in the reference genome sequence, an error generated in the library, misalignment or base-calling error. Steps can be taken at every stage in the variant detection process to avoid such errors:

7.1.1.1 Avoiding errors generated in the library

The PCR amplification step in library preparation can result in incorrect DNA fragments in the final library, which show up as mismatches in the alignment, falsely indicating genetic variation in the sample. Duplicate DNA fragments may falsely manifest as high read depth/coverage. PCR-free protocols do exist (such as Illumina's TruSeq DNA PCR-Free Sample Preparation Kits), but they require large amounts of input DNA, are very sensitive to variations in DNA quality and purity, and do not necessarily guarantee unbiased sequencing data; mean coverage, for example, was shown to be equal when using PCR-amplified and PCR-free protocols [965]. With PCR-amplified libraries, PCR duplicates are commonly removed after alignment, using programmes such as Picard (SourceForge, Dice Holdings, Inc., <http://picard.sourceforge.net>) [966].

7.1.1.2 Avoiding base-calling error

After sequencing is complete, the quality of raw reads must be evaluated. With many base-calling algorithms, uncertainty can be quantified, as each nucleotide base call is accompanied by a statistical measure of uncertainty or Phred quality score – a Phred score of 20 corresponds to a 1% error rate in base-calling [967]. The researcher may filter out bases to include only those of high confidence. Poor-quality tails of reads (bases at the 3' or 5' end of a sequencing read) may be trimmed off, although the importance of this filter is debatable: it has been shown to reduce expected error rates (the average error probability of each base assigned as quality scores) from 7.09% to 0.16% and from 16.43% to 0.23% [968], but others found that trimming led to more alignment artifacts and thus caused more false-positive variant discovery than when using raw reads [969]. The authors stated that ends of reads may be helpful for read mapping. Reads may also be filtered according to other parameters, such as reads with at least one uncalled base (N content) or GC bias [966, 968]. It has been reported that, with Illumina data, erroneous base calls are

frequently preceded by base G [970] and A-C base conversions are the most frequent [970, 971].

7.1.1.3 Avoiding alignment (or mapping) error

When aligning reads to a reference genome, either hash-based alignment algorithms, such as MAQ [972], Novoalign (<http://www.novocraft.com/main/index.php>) and Stampy [973], or Burrows-Wheeler transform-based algorithms, such as Bowtie [974], SOAP2 [975] and BWA [440], are used. Hash-based alignment algorithms are based on a hash table – a common data structure that may be created using either the reference genome or the set of sequencing reads [976]. Either the reference genome is used to scan the hash table of sequencing reads, or the sequencing reads are used to scan the hash table of the reference genome. The reads are then associated with the region of the genome where they are most likely to align. With Burrows-Wheeler transform-based algorithms, an index of the reference genome assembly is first created by modifying the sequence order of the reference genome so that sequences that exist multiple times now appear next to each other and then the final index is created. This facilitates rapid read placement along the genome. Hash-based alignment algorithms, although slower, are more accurate than Burrows-Wheeler transform-based aligners and the former are recommended, especially for short reads [967, 976].

Alignment is more difficult for regions with higher levels of diversity between the reference genome and the sequenced genome, as more mismatches between the sequencing reads and reference genome must be allowed, and also for shorter reads (which, from next-generation sequencing, are already shorter than reads from Sanger sequencing), as there may be several places in the reference genome from which they could have been read. This difficulty can at least partly be overcome by using longer reads and paired-end reads, where long fragments are taken from the donor genome and a number of bases are

sequenced from both ends of the fragment, rather than from only one end (single-end reads) [973, 977]. As the distance between each paired read is known, alignment algorithms can use this information to map the reads more precisely, which is particularly important for repetitive regions of the genome [978].

7.1.1.4 Calling variants

Once high-confidence base calls are aligned to position, variant calling is conducted. This is often separated into two steps: genotype assignment and variant identification [979]. It is determined at which positions at least one of the bases differs from the reference sequence and then, for positions in which a variant has already been called, the genotype can be determined. GATK [980] and SAMTools [981] are two popular tools used for variant identification.

Genotype-assignment is conducted using probabilistic methods, where the probability of a particular genotype at a particular site is calculated, assuming independence among reads, and the genotype with the highest probability is generally chosen. A measure of statistical uncertainty for the genotype call is also provided. The prior probability for each genotype, which must be estimated in order to produce posterior probabilities for the genotypes, may assign equal probability to all genotypes, or external information (for example, from the reference sequence, SNP databases or an available population sample) may cause more weight to be given to a particular genotype over others. It has been recommended that prior probabilities are calculated by analysing multiple individuals at the same time (multi-sample variant/genotyping calling), as allele frequencies (either from the dataset in question or from population databases) can be used to estimate how likely the presence of a particular genotype is for an individual [967]. The probability that all genotypes are homozygous for the reference type can be estimated, giving both an SNV call and an associated measure of confidence. LD patterns at nearby sites can also be used to improve

genotype-calling accuracy [967], although this may not be of much use for rare variants that are not present in reference databases.

Most variant-calling software focuses on detecting SNVs, but indels can also be detected from sequencing data, either by gaps in the read mapper alignments, *de novo* assembly of short reads and comparison of contiguous, overlapping sequence reads (contigs) to a reference sequence, or from lists of indels available in databases, such as dbSNP [987].

Detection of indels can be difficult, as they occur at a lower rate than SNVs, the position of an indel with respect to the reference genome is often ambiguous (they may not be uniquely positioned onto the reference, especially if they are located in repetitive regions of the genome) and there is much variation in fluorescence signal intensity for a specific homopolymer length [988], often resulting in high error rates in indel calls. Error rates are affected by indel frequency (extremely high indel frequencies may decrease sensitivity of indel-calling algorithms), read length (increased read length is associated with improved performance of indel-calling algorithms), indel size (longer indel length is associated with lower positive predictive value in some softwares), and read depth (increased coverage is associated with increased indel-calling sensitivity) [989, 990]. It is common to realign reads around indels, as misalignment of reads where the first or last few bases overlap an indel may lead to false SNV calls – ungapped alignments (those that do not allow gaps in alignment) optimise local alignment by minimising the number of mismatched bases [986, 990-992]. Use of tools that allow gaps in alignments can reduce error rates in SNV detection [990]. It has been recommended to call SNVs after realignment of reads by the indel-calling algorithm [987].

Detection of larger structural variant detection from exome sequencing data is a challenge, as it is sparser than genome sequencing data and there is more variation in read depth.

Variants >15bp have rarely been reported in exome studies [996]. Some methods have

been developed to detect CNVs from exome sequencing data, such as CoNIFER [997], which has an estimated sensitivity of approximately 76% for rare CNVs containing three or more exons. The authors acknowledge that some exonic CNVs are likely to be missed and state that 'owing to the significantly lower probe density and the targeted nature of exome-capture platforms, our algorithm cannot fully replace currently available high-density array-comparative genome hybridisation or SNP microarray platforms if genome-wide detection of CNVs is the goal.' Another algorithm, Splitread [996], uses a split-read approach whereby clusters of paired-end reads with one end that maps to the reference genome but not at the other end (because it crosses a breakpoint) are searched for. The unmapped ends are broken down into sequences either of equal length (balanced splits) or unequal length (unbalanced splits) and clusters of split reads are searched for, using the balanced splits as seeds. The authors claim that Splitread can detect insertions and deletions of any size, whilst acknowledging that accurate characterisation of insertions is influenced by read length and that some genetic variation was still missed. Concordance between multiple algorithms was recommended, as not all calls predicted by Splitread were also predicted by other methods, and a substantial fraction of calls were unique to a particular method.

One can increase the accuracy of SNV/indel/CNV calling by increasing depth of coverage – the number of times a nucleotide is read during the sequencing process. With low-coverage sequencing (an average of $<5\times$ per site per individual), there is high probability that only one of the two chromosomes of a diploid individual has been sampled at a specified site [967]. As sequence coverage increases, genotype likelihoods should peak around the true genotype, as more sets of reads contain a particular allele [998]. When coverage is low and there may be no reads covering a site, LD patterns in nearby sites may be taken advantage of to infer genotypes [998]. With indel-calling, some algorithms have

been shown to reach a plateau of true indel calls when coverage is >30, whereas others showed an increase in true indel calls as average coverage increased [989]. Higher coverage provides higher resolution of CNV break point detection, giving higher accuracy [999, 1000]. Of course, reducing uncertainty by increasing coverage (to either medium coverage [5-20x] or deep sequencing [>20x]) is often prohibited by financial constraints and, even with average higher coverage, there is still variation in coverage across regions [1001]. Researchers may rely upon the proportion of targeted base-pairs that are covered at (or greater than) a particular depth (for example, 20x). Some target regions may have a lower coverage due to, for example, an increased proportion of G and C bases in a region [1002], or higher coverage due to, for example, similarity between intergenic and exonic regions, causing non-specific binding during the capture process [1003]. It has been shown that, in association studies, a greater increase in power is gained by sequencing more individuals at a lower depth than by sequencing fewer individuals at high depth [1004].

Even if parameters such as coverage are consistent, variant detection still depends on the algorithm used. Discrepancies have been observed between different algorithms both for SNV [1005-1007] and indel [990, 1005, 1006, 1008] detection, which suggests high false negative and false positive rates of these algorithms. Concordance between indel-calling algorithms (26.8%) has been shown to be lower than with SNV-calling algorithms (57.4%) [1005]. Indels of low frequency may be less likely to be called by multiple algorithms [990]. Just as for larger structural variations, concordance between multiple algorithms is recommended. This approach is supported by findings that indel events called by two or more methods had the highest concordance with dbSNP and 1000 Genomes Project calls [996].

7.1.1.5 Variant filtering

Post-calling, variants can be further filtered according to various parameters in order to obtain a set of high-confidence calls. Such parameters include read depth, genotype quality score, differences in quality scores for reference and alternative alleles, allelic balance ratio, base quality score, mapping quality score, deviations from HWE, and proximity to indels [967, 986, 1003]. Filtering can be undertaken with variant alignment software, such as GATK and SAMTools, or a programme designed for manipulation of variant calling files (vcf), Vcftools (SourceForge, Dice Holdings, Inc., <http://vcftools.sourceforge.net>). As for the filtering steps that can be conducted prior to variant-detection, there is no gold standard as to which thresholds to adopt. The most common algorithms have default settings and variants can also be filtered according to other parameters at the researcher's discretion.

In the output vcf files, a QUAL score is given for each variant - a measure of how confident the algorithm is that the SNV is true. Researchers may filter variants based on this QUAL score, as well as on other scores such as read depth, although recommendations vary.

GATK does not recommend applying a hard filter for read depth with exome sequencing data as, with capture data, the relationship between misalignments and depth is not clear.

It has been reported that a QUAL score of approximately 20 and a read depth ≥ 5 can generally be applied to give a high validation rate when using variants called by SAMtools:mpileup [1011]; these thresholds were determined by comparing accuracy versus recall with different combinations of score cut-offs, although the authors acknowledge that thresholds may be dataset-specific. The authors recommended checking allele/strand bias but applied no filtering rules based on allele/strand balance (reads supporting reference allele forward, reference allele reverse, alternate allele forward, and alternative allele reverse) in their dataset, as neither accuracy nor recall improved substantially. Others have excluded SNVs with read depth < 10 and those located outside of exome-capture regions [1005], or SNVs with a quality score < 20 and indels < 50 [1011].

The quality of SNV calls can be evaluated by checking performance metrics, such as transition to transversion (Ti/Tv) ratio and genotype quality score. Filtering criteria can then be adjusted to an optimum level, whereby both high genotype consistency rates and Ti/Tv ratios close to expected values are achieved: Ti/Tv ratios of approximately 3.0 (3.5 for known variants and 3.0 for novel variants) are expected for whole exome datasets and approximately 2.0 for whole genome datasets (2.10 for known variants and 2.07 for novel variants) [980, 1012, 1013], although the ratio for synonymous variants is higher than for non-synonymous variants [945, 1014]. In one study, the two most effective thresholds affecting sequencing data genotype quality were found to be genotype quality score ≥ 20 and depth ≥ 5 – these filters resulted in both high genotype consistency rates and Ti/Tv ratios close to expected values [1003]. Accuracy of an algorithm can also be estimated by comparison of results with calls from SNP genotyping platforms, array-comparative genome hybridisation data or Sanger sequencing verification. The proportion of calls reported in dbSNP provides a rough measure of accuracy [980], as do heterozygosity rates, as sequencing can easily be mistaken for heterozygous sites, especially where coverage is low [1015]. A visual examination of reads – checking coverage and quality of local alignment around the site of variation – has also been recommended for the identification of false positives [1011].

7.1.2 Strategies for identifying susceptibility variants or genes in exome sequencing data

Another challenge with next-generation sequencing data is how to identify variants or genes that are associated with the phenotypic trait amongst a plethora of non-pathogenic variants/genes (assuming, of course, that sequencing error has been eliminated as much as possible).

One approach is to prioritise variants based on novelty, or rather lack of presence in public databases, such as dbSNP or 1000 Genomes Project, based on the premise that common variants are less likely to be disease-causing. This does exclude a high proportion of variants, leaving only around 2% of SNVs identified by exome sequencing [935], but it also assumes that such databases contain no pathogenic variants, which is not the case. In family studies, *de novo* mutations may be prioritised, although these are more likely to be relevant in severe early-onset disorders where transmission to subsequent generations is often prevented – they are under strong negative selection [943].

Variants can be filtered solely by MAF, based on the premise that deleterious variants are subject to purifying selection and tend to show high sequence conservation. The MAF threshold for ‘rare’ variants, however, is arbitrary – normally defined as either 1% or 0.5% in publications, but variants with MAFs as high as 5% have been described as ‘rare’ [868, 869], rather than ‘low frequency’. Excluding variants with a very low MAF in the general population may erroneously exclude causal variants. For example, if a disorder is recessive, the alternative allele may be present at a higher frequency in the general population, as its deleterious effects would be masked in the heterozygous form, compared to a dominant disorder. This is the case with the *CFTR* gene for cystic fibrosis.

As outlined in Chapter 6, filtering variants based on class – those that are most likely to be deleterious, such as non-synonymous or loss-of function variants – is a common approach, but does not account for variants that indirectly alter protein-coding sequences. Moreover, high levels of functional genetic variation are present in the exomes of healthy individuals, and around 100 variants are thought to be loss-of-function [798, 815, 935, 947, 992, 1016]. Predicting deleteriousness based on bioinformatic tools is an alternative method, although there is often discrepancy between results from different algorithms and most are restricted to non-synonymous variants. PhastCons, PhyloP and GERP++ do give scores for

synonymous variants, but if one is looking for concordance with other algorithms that account for other factors besides evolutionary conservation, all variants apart from non-synonymous SNVs will be most probably be filtered out. Filtering according to algorithms may lead to erroneous exclusion of susceptibility variants. For example, even if most SNPs associated with Mendelian disease are located in highly conserved regions, this may not necessarily be the case for those associated with complex traits [1017] and so filtering variants according to PhastCons, PhyloP and GERP++ scores, for example, would not be appropriate. Rather than using these algorithms to filter variants in the first place, it may be best to use them post-analysis, to strengthen existing evidence of pathogenicity/association.

As well as prioritising variants based on putative functional consequence, variants can also be pooled into genes. This may be particularly important when searching for the genetic contribution to complex traits, where power to detect an association is often limited by low penetrance, small effect size, locus heterogeneity, variable expressivity and pleiotropy [1018]. It does not have to be the same variant that is over-represented in one responder/disease group, as long as the variants all affect the same gene and are likely to be of functional consequence [863]. As exonic variants can typically be clearly assigned to a gene, unlike with GWAS, where variants are often intergenic, gene-based tests are commonly used with sequencing data.

Association tests may be more suitable than discrete filtering for complex traits, where variants that influence disease risk or a phenotypic trait may not be fully penetrant, mutations in different genes may lead to the same phenotype, or more complex phenotypes may be explained by the presence of multiple variants [943]. Gene-based association tests are commonly performed with whole genome or exome sequencing data to increase power to detect genotypic-phenotypic associations [806, 1019]; these data are

particularly suitable for such analyses as one is not limited by the content of an array – theoretically, every variant is captured. Unless rare variants have a large effect on the phenotype, there is limited power to detect an association with a single rare variant, even with large sample sizes – it is expected that over 10,000 exomes will be required to achieve sufficient statistical power to robustly detect associations of rare variation with complex traits [944]. In gene-based association tests, group frequency or distribution, as opposed to single variant frequency or distribution, is compared between cases and controls. In the absence of pedigrees, individuals at extreme ends of the phenotypic spectrum are often selected, based on the premise that these individuals are genetically more informative and that the frequency of disease/trait-associated alleles are enriched in one or both ‘extreme’ groups [855, 856].

There is a caveat with gene-based tests: not all variants within a gene may affect its function and those that do affect gene function may affect it in different directions [963]. Examples of genes where different variants have opposite effects on the phenotype include *PCSK9*, in which loss-of-function variants are associated with reduction in mean LDL cholesterol and gain-of-function variants are associated with decreased LDL receptors and hypercholesterolaemia [1020], and *APOB*, in which rare missense mutations can result in hypercholesterolaemia and hypocholesterolaemia [1021]. Certain gene-based tests, such as the c-alpha test [660] (a test of dispersion), account for the fact that some variants may be protective and others deleterious with regards to the phenotype in question. Other tests assume that the effects of all variants within a gene are in the same direction: damaging with regards the phenotypic trait.

Variant prioritisation, for example by variant class, may also be performed so that only putative functional variants are included in gene-based analyses, although recent studies have highlighted the importance of not restricting analyses solely to variants in coding

regions. For example, Harismendy et al. sequenced individuals with extreme BMIs and used a collapsed marker test to identify rare variants associated with extreme obesity [1026]. Most of the associated variants were located in regulatory elements, either close to the gene promoter or in transcriptional enhancers. Furthermore, none of the associated variants in the aforementioned study were present in evolutionarily conserved sequences, as determined by PhyloP score, which highlights the dangers of missing true genotypic-phenotypic interactions when filtering variants based on prediction algorithms.

The aforementioned methods, or a combination of them, have been used to increase power in association studies and pinpoint candidate genes. The 'novelty' approach was adopted in the first publication to show, as proof-of-concept, that exome sequencing can identify the cause of Mendelian disorders [874]. The authors searched for genes with non-synonymous, splice acceptor and donor site variants and short coding indels, not present in public databases, present in all cases (individuals with Freeman-Sheldon syndrome) but no controls (individuals from the HapMap and Human Genome Structural Variation projects); one variant had to be present in the case of a dominant disorder, or two (in the same gene) in the case of a recessive disorder. After such filtering, the only gene remaining was *MYH3*. Filtering variants by those predicted to be damaging according to PolyPhen also served as an effective filter, again narrowing down the candidate gene to *MYH3*, although the previous filters (only including genes containing at least one rare coding indel, non-synonymous or splice-site variant, and excluding variants present in dbSNP [common variants] and in any controls) were sufficient. Others have prioritised genes in which all affected individuals had novel variants at different genomic positions (making it more likely that the variants are causative and not simply unidentified SNPs) [1027], or genes with at least one previously unobserved heterozygous non-synonymous or splice site substitution or a coding indel in all affected individuals [1028]. This approach assumes not only that the

variant is rare and likely to be previously unidentified, but also that the variants are fully penetrant. A similar strategy has been previously adopted that allows variants to be present in a proportion of cases, not necessarily all cases, which allows for a degree of genetic heterogeneity and/or missing data [1029].

Many family-based exome studies have targeted *de novo* mutations, in particular those thought to be deleterious, based either on variant class or functional prediction algorithms. O’Roak et al. found there to be a higher rate of non-synonymous *de novo* variants in trios with autism spectrum disorder than expected [961, 1030]. Variants were classed as severe if they were truncating, missense with Grantham score ≥ 50 and GERP score ≥ 3 or only Grantham score ≥ 85 , or deleted a highly conserved amino acid [1030]. Sander et al. narrowed down candidate genes by looking for ≥ 2 ‘disruptive’ *de novo* mutations in the same brain-expressed gene in unrelated individuals, as this was unlikely to occur by chance, due to the very low rates for *de novo* nonsense and splice-site mutations [1031]. The authors found that functional prediction algorithms did not provide any more information with regard to disease risk, compared to classifying variants based on class (non-synonymous versus synonymous), and then further filtered to nonsense and splice site mutations. In such studies, it must be considered that the probability that a gene contains multiple gene-disrupting *de novo* mutations by chance increases with sample size, and that there is a higher probability of observing multiple missense variants, compared to gene-disrupting variants, as the mutation rate generating missense is approximately 20 times higher than that generating gene-disrupting variants [1032].

Which strategies to adopt depends on practical considerations – whether one has samples from a pedigree or unrelated probands, how many samples can be sequenced with the available budget, and various assumptions: whether the phenotypic trait is thought to be due to *de novo* or inherited variants (for family-based studies) and the assumed mode of

inheritance of a trait, the extent of locus heterogeneity for a trait (whether or not the trait is caused by variants in genes at different chromosomal loci), whether only rare variants or both rare and common variants are expected to contribute to the phenotypic trait, or whether protective and risk variants are expected [935, 1033].

7.1.3 Pathway analysis

As with array data, grouping variants based on their role(s) in a biological pathway, such as those defined by PANTHER ([671], <http://www.pantherdb.org/pathway>) or KEGG ([1034], <http://www.genome.jp/kegg/pathway.html>), can increase power if a trait is affected by multiple causal genes in a pathway. One may apply the same principles as with GWAS, for example, by assigning SNPs to a gene, taking the SNP with the lowest p-value to be the p-value for that particular gene, and determining whether a particular pathway or biological function is over-represented in a set of genes with p-values below a specified threshold, compared to what would be expected by chance. One may also use results from gene-based tests, although results will be affected by the assumptions of the original analyses (for example, gene-based tests that assume all variants have a uni-directional effect will generate different p-values to tests that allow for bi-directional effects). The limitations of this competitive approach are the same as with genotyping data in that not all SNPs/genes are represented in the list of nominally significant genes. Furthermore, a major difference between array genotyping and sequencing is that only one or a few variants per gene are represented on arrays, whereas with exome sequencing, all exonic variants in a gene are included. If several variants within a gene are of moderate effect size, taking the SNP with the lowest p-value may result in a loss of information [666]. A self-contained analysis, where one extends gene-based analyses by assigning variants to a pathway or a particular GO term, instead of just to a gene, to test the null hypothesis that no pathway is associated with the phenotypic trait, avoids this loss of information.

7.1.4 Population stratification

Whichever strategy is adopted, for studies involving unrelated individuals, population structure requires consideration, as rare variants, which are typically the result of recent mutations, are likely to segregate in a population-specific manner [875, 945]. For example, a deficiency of rare alleles has been observed in Europeans and East Asians compared with expectations for a constant-sized population [1035]; West Africans had fewer rare alleles than expected. The former was attributed to the 'out-of-Africa bottleneck' or decline in population size and consequential reduction in genetic variation as people migrated from Africa, and the latter was attributed to an expansion in African population. East Asians were found to have even fewer rare alleles than Europeans as East Asians had greater genetic drift (varying frequency of a gene variant in a population over time due to chance) throughout history. Controlling for potential population substructure is particularly important for exome sequencing studies, where discovering rare variation is often the goal. In an association study, if a higher proportion of cases had African ancestry compared to controls, one would expect to observe an excess of rare variants in cases, not due to any influence the variants have on the phenotypic trait, but because Africans harbour more rare variants than non-Africans [944].

In contrast to GWAS, methods for achieving this with sequencing data are not fully established and it is unknown whether the methods traditionally used for detecting and controlling for population stratification in studies with common variants are suitable for studies that also contain rare variants [1036]. Ideally, as with GWAS, cases and controls would be matched for population proportions and disease risks in subpopulations. However, this is not always feasible, particularly when matching is based on indices such as geographical proximity or self-reported ethnicity and thus subpopulations may remain

unidentified; for example, participants may not know about their genetic admixture. A variety of alternative approaches have been used thus far.

Principal component analysis: In an analysis of rare variants from an exome data set, principal component analysis (using three vectors, obtained from common [MAF>0.05] non-synonymous SNPs) reduced Type I error due to population stratification, but it did not improve the power to detect causal variants [1037]. Ten principal components were incorporated in logistic regression based analyses of exome sequencing data, which mitigated spurious association rates due to population stratification, although power was reduced in every case [1038]. In another study, which used exon pilot data from the 1000 Genomes Project, with subjects from seven populations, the inclusion of principal components as covariates in rare variant aggregation tests tended to (but not consistently) reduce Type I error rates [1039]. The top ten principal components, based on rare and common variants, were used as covariates to control for population stratification in a sequencing dataset [1040]. This was effective with a small number of sub-populations, although the authors note that this approach may not be sufficient in other scenarios. Although inclusion of a large number of principal components will remove virtually all stratification, it is unknown how many are required and the more components included, the lower the power to detect true associations [1041].

Permutation: Kiezun and coauthors claimed to effectively control Type I error in association tests conducted on exome sequencing data by ‘modifying the permutation scheme to account for subpopulations’: case-control status was permuted within each population, assuming known population labels [944].

Genomic control λ : using simulated genome-wide genotype data, it has been shown that, with gene-based tests such as the burden and c-alpha test, the inflation caused by

population stratification varies across genes, depending on their size and on the MAFs of its variants [1036]. Therefore, λ_{GC} , which inflates the association-test statistic by a common factor, is not suitable for gene-based association tests. It was showed that, when there are two sub-populations, using the top few vectors from principal component analysis, constructed from either common or rare variants, controlled for population stratification, but this was not achieved when ten sub-populations were present.

7.1.5 Interpretation and utility of results

The final challenge regarding sequencing data is the interpretation and subsequent usefulness of obtained results. Assuming that mapping and variant-calling errors are controlled for, finding variants is not necessarily the greatest problem, but rather how to identify the particular mutation relevant to the disease or phenotypic trait and how exactly it exerts an influence. Excluding all non-loss-of-function variants may not necessarily be the answer, as a typical exome harbours approximately 100 loss-of-function variants [815]. Multiple observations, either of the same variant (which is unlikely in the case of complex disorders) or of independent variants in the same gene either in cases or controls, increase confidence when classifying particular alleles or genes as risk-conferring, even if association test results do not reach statistical significance, for example due to a small sample size. Once one or a small group of candidate genes have been identified, targeted sequencing of these genes, for example by Sanger sequencing, can help verify the association of specific variants with the trait in question [1033]. Just as with genetic association analyses based on genotyping arrays, functional studies can help interpret association signals, providing insight into the mechanisms behind disease/trait risk [945].

As for the clinical utility of sequencing data, it is clear that, at least with Mendelian disorders, it can facilitate diagnosis, in particular when clinical presentation or biochemical parameters may not be indicative of the disorder. This may or may not alter clinical

management. The clinical application of genome/exome sequencing for complex diseases is not yet fully established [935], although it is likely to provide insight into the allelic and/or locus heterogeneity of such diseases and how this relates to phenotypic variation.

It is hoped that whole exome sequencing will identify one or more genes/pathways that influence KD, or help elucidate some of the mechanisms of action of dietary treatment for epilepsy.

7.2 Methods

7.2.1 Phenotypic data

As in Chapter 6, the extreme trait phenotyping strategy was adopted in order to identify those individuals most likely to have highly penetrant rare mutations that affect KD response. Funds were available for sequencing of 45 samples.

18 extreme responders and 27 extreme non-responders who had enough DNA available for whole exome sequencing were selected. The 18 responders had diet response data for at least 12 months and had maintained $\geq 75\%$ seizure reduction at all follow-up points. The 27 non-responders reported an increase or no change in seizure frequency and weaned off the diet either prior to the 3-month point or approximately at the 3-month point due to a lack of response or adverse response.

The following cases were not considered for inclusion in the extreme non-responder category for exome sequencing:

- individuals who weaned off the diet due to lack of tolerability or compliance.
- individuals who experienced a dramatic reduction in seizure frequency soon after starting the diet but whose seizure frequency subsequently increased so that, by the time they

reached the 3-month clinic appointment, seizure frequency was no different to prior to starting the diet.

- individuals who had no change in seizure frequency but did experience other seizure-related benefits.

- individuals who reported any non-seizure-related benefits and who continued following the diet due to these benefits.

A brief outline of the chosen 45 individuals' experience with the KD is given in Table 7.1.

Table 7.1: Description of Ketogenic diet response in 45 samples who had whole exome sequencing

Extreme responders	Response to KD
Case 1	>75% reduction in seizure frequency for 2 years. Seizure control was then variable and the patient was weaned off the diet 5 years after commencing the KD as further manipulations had not led to improved seizure control. There was no deterioration in seizure control.
Case 2	>75% reduction in seizure frequency for the 2.5 years that the patient followed the diet. The patient then 'weaned' onto a low glycaemic-index regime (not supervised by a dietitian) and seizure control has remained variable but not as poor as prior to starting the KD.
Case 3	>75% reduction in seizure frequency for the two years that the patient followed the diet. During the last few months on the diet, the patient showed signs of abdominal pain, which resolved when cow's milk was eliminated, although episodes of seemingly generalised pain later returned. No change in seizure control when weaned off the diet and episodes of pain gradually disappeared.
Case 4	'Dramatic' >75% reduction in seizure frequency for the 18 months that the patient followed the diet. After weaning off the diet, the patient initially experienced more frequent clusters of epileptic spasms/tonic seizures but seizure control subsequently improved.
Case 5	>75% reduction in seizure frequency for the 18 months that the patient followed the diet. The diet was then weaned due to behavioural problems – obsessions with food – although the patient remained on a gluten-free diet. There was no deterioration in seizure control.
Case 6	Consistently >90% reduction in seizure frequency over the 4 years when the patient followed the diet. Diet was weaned due a loss of efficacy.
Case 7	Consistently >90% reduction in seizure frequency over the 8 years the patient has followed the diet. The patient is essentially seizure-free except during attempted weaning of diet when seizures returned.
Case 8	>75% reduction in seizure frequency during first three months on the diet. Subsequently became seizure-free with recurrence of seizures just before the 24-month point. Seizure-control improved (not quite seizure-free). The patient was weaned off the diet 3.5 years after starting, with no deterioration in seizure control.
Case 9	Became seizure-free during the first three months after starting the diet. Remained almost seizure-free for two years. When attempting to wean the diet, seizures recurred. The patient went back on the diet, with some improvements, and is still following

	the diet.
Case 10	A 'dramatic' improvement in seizure control, noted 3-4 days after starting the diet. This improvement (almost seizure-free) was maintained for 2.5 years, when seizure frequency increased. The patient was weaned from the diet but further deteriorated. The patient was not put back on the diet (as of Oct 2013) due to problems with digestive system/bowel habits.
Case 11	Almost seizure-free when following the diet for three years. Seizure control then became variable but still better than prior to starting the diet. Died 4.5 years after starting the diet
Case 12	Became seizure-free during the first month after starting the diet. The patient had been following the diet for 2 years 3 months and remained seizure-free, with the exception of one isolated incidence of seizure exacerbation, which subsequently came under control, and one seizure in Feb 2013 associated with febrile illness. The patient was weaned off the diet to improve oral intake and reinitiated the diet (and AEDs were increased) in September 2013 due to a hospital admission with infection and breakthrough seizures, which lead to an improvement. The patient is currently weaning off the diet again.
Case 13	>75% reduction in seizure frequency in first three months on the diet. This further improved to >90% reduction, which is still the case 10 years later (still on diet). The patient has periods of seizure-freedom.
Case 14	Has been following the diet for over two years and still has a 90% reduction in seizure frequency (achieved in first three months on diet).
Case 15	Became seizure-free three weeks after starting the diet. The patient remains seizure-free 24-months later (still on diet). Subsequently diagnosed with GLUT-1 deficiency.
Case 16	70-80% reduction in seizures during first three months on the diet. The patient has now been following the diet for 18 months and approximately 80% reduction in seizure frequency has been consistently achieved over this time, with the exception of periods of illness when frequency increases.
Case 17	Consistently >90% reduction in seizure frequency over the 3 years the patient has followed the diet.
Case 18	Became seizure-free during the first three months on the diet and the patient currently remains seizure-free 2.5 years later (still on diet).

Extreme non-responders	Response to KD
Case 1	No change in seizures. Weaned off diet before the 3 month-point
Case 2	No change in seizure frequency. The patient experienced longer events.
Case 3	Eyelid myoclonia increased in frequency and severity/more prolonged; no change in vacant spells
Case 4	No change in seizures. Weaned off diet before the 3 month-point
Case 5	No change in seizures.
Case 6	No change in seizures. Weaned off diet before the 3 month-point. Increased brightness, awareness and interaction, but thought to be due to reduction in clobazam dose.
Case 7	No change in seizures
Case 8	No change in seizures. The only reported benefit was an improvement in bowel habits
Case 9	No change in seizures
Case 10	No change in seizures. Weaned off diet before the 3 month-point. Died a few months later.
Case 11	No benefit from diet - possible seizure increase
Case 12	No change in seizures. Weaned off diet before the 3 month-point
Case 13	No benefit from diet - possible seizure increase. Weaned off diet before the 3 month-point
Case 14	No benefit from diet - possible seizure increase.
Case 15	No change in seizures.
Case 16	No change in seizures for almost 3 months. The patient then had an episode of status epilepticus, the first such episode for five years. The patient was weaned off the diet.

Case 17	Clinic letters indicate no change in seizures. Seizure diary indicates an increase in seizure frequency.
Case 18	Increase in frequency of absence seizures at 3-month point.
Case 19	No benefit from diet - possible seizure increase.
Case 20	Increase in seizure frequency. Weaned off diet before the 3 month-point. Parents stated that seizures are 'currently the worst ever and have been escalating since the patient started the diet'.
Case 21	Increase in seizure frequency.
Case 22	Patient stated that there were no improvements in seizure frequency or pattern. Seizure diary showed an increase in seizure frequency
Case 23	Sudden increase in seizure frequency. As soon as the patient went into ketosis, the patient deteriorated rapidly. Seizure clusters could not be stopped and the patient was admitted to hospital.
Case 24	Increase in seizure frequency but seizures slightly shorter.
Case 25	Increase in seizure frequency. The patient started having seizures from wakefulness during the day, which hadn't happened in a very long time. These evolved into a nocturnal cluster requiring emergency medication. Weaned off diet before the 3 month-point.
Case 26	Increase in seizure frequency. Weaned off diet before the 3 month-point.
Case 27	'Dramatic' increase in seizure frequency. Weaned off diet before the 3 month-point.

7.2.2 Genotypic data

DNA for the 45 extreme responders and non-responders were measured with NanoDrop ND-1000 Spectrophotometer (Thermo Fisher Scientific Inc., USA) and diluted to 60µL at 30ng/µL by the researcher. Samples were given to Miss Deborah Hughes at UCL Institute of Neurology Neurogenetics Unit in a random order (so that all responders were not sequenced before non-responders, or vice versa) for whole exome sequencing. Samples were sequenced in two batches, although data from several runs were combined for six samples.

The sample library was optimised using the Nextera DNA Sample Preparation Kit (Illumina, Inc., San Diego, USA) and targeted regions were captured with the Nextera Exome Enrichment Kit. Illumina HiSeq® 2500 was used as the platform for sequencing. 8GB per sample was sequenced.

The following steps were completed by Dr Alan Pittman at UCL Institute of Neurology Neurogenetics Unit:

- i) Alignment of sequencing reads to the reference genome sequence Genome Reference Consortium Human Build 37 with Novoalign (Novocraft Technologies Sdn Bhd, Malaysia, <http://www.novocraft.com/main/index.php>). Novoalign is a hash-based alignment algorithm, which, although slower, is more accurate than Burrows-Wheeler transform-based aligners [967, 976]. Novoalign allows gapped alignments and is thus more sensitive to indel detection: in one study, approximately 90% of indels of a particular size were correctly detected in reads aligned by Novoalign [990]; <50% were detected in BWA alignments.

- ii) Removal of PCR duplicates with Picard (SourceForge, Dice Holdings, Inc., <http://picard.sourceforge.net>), to leave only unique reads.
- iii) Calling of variants and creation of .vcf files from BAM alignment files using SAMTools (SourceForge, Dice Holdings, Inc., <http://samtools.sourceforge.net>, [981]).

The following commands were used to call quality-controlled filtered variants:

```
mpileup -q 20 -L 400 -d 400
```

-q 20 = skip alignments with read mapping quality smaller than 20

-L 400 = skip indel calling if the average per-sample depth is above 400

-d 400 = at a position, read maximally 400 reads per input BAM

Variants were then further filtered using SAMTools default settings. Variants with a QUAL score of <18 were also removed.

SAMTools sets the prior probability for a heterozygous SNV at 0.001 for novel variants, and 0.2 for known SNVs [1005] and has a mean sensitivity of 94.47 and mean specificity of 99.59. SNV calling by SAMTools has been called conservative, as a larger number of false negatives were observed, compared to the programmes 'Variant Caller with Multinomial probabilistic Model' and GATK [1006]. Indels are called from the pileup of reads at every position along the reference sequence. SAMTools has been shown to identify a larger number of indels compared to the other programmes and was found to have a high false positive rate; SAMTools also identified a higher number of common indels compared to the other two methods, indicating a lower false negative rate [1006].

- iv) Annotation of variants with ANNOVAR

(<http://www.openbioinformatics.org/annovar> [846]).

7.2.2.1 Quality control

7.2.2.2 Per-variant quality control

In addition to the criteria applied when calling SNVs and indels, performance metrics were calculated to see if the quality of the data could be improved by applying further thresholds. Vcf files were manipulated and merged by the researcher, using Vcftools (v0.1_11, SourceForge, Dice Holdings, Inc., <http://vcftools.sourceforge.net>), to create one file including all exonic variants for all individuals (used for performance metrics) and one including all exonic and splicing variants for all individuals (used for association analyses). These merged vcf files were imported into PLINK/SEQ (<https://atgu.mgh.harvard.edu/plinkseq>).

The Ti/Tv ratio, proportion of variants present in public databases (such as dbSNP, 1000 Genomes Project [April 2012 release], or NHLBI [data release ESP6500]), number of singletons, and mean genotype quality score were calculated for each individual (PLINK/SEQ command: *i-stats*), firstly when no additional quality control thresholds were applied, and then when adjusting certain thresholds (for example, excluding genotypes with a quality score [GQ - Phred-scaled probability that the called genotype is wrong] <10 or <20). Such statistics are useful as broad indicators of the quality of the data set [944]. The Ti/Tv ratio is a useful diagnostic metric because transitions (A<-> G and C <-> T) naturally occur more often than transversions (A <-> C, A <-> T, G <-> C or G <-> T); for exons present in the Consensus Coding Sequence Project database, the average Ti/Tv ratio should be close to 3:1 [1042]. The percentage of variants found in public databases is used to obtain an approximate false-positive rate of variant calling. There is no standard definition of the proportion of variants that should be present in public databases

(previously-reported proportions vary widely: in one study, 85-88% of identified variants were found to be present in dbSNP [1043], depending on the platform used for sequencing; in another, an average of 92% of variants were already annotated in dbSNP (v129) [874]), but individuals with a lower proportion compared to others in the dataset may indicate an increased false positive discovery rate. It should be noted that the proportion of novel variants varies between regions with different ancestries; individuals of Native American ancestry have a higher proportion of novel variants compared with those of European ancestry, and individuals of African ancestry have the highest rates of novelty [814]. Random errors in sequencing are likely to be limited to a single sequence within the population sample and so an elevated number of singletons could indicate that errors are present [1044]; it has been stated that the proportion of singletons should be below 10% [1012, 1042], although the 'normal' number of singletons will depend on individual genetic diversity.

Additional quality control filtering made little difference to the performance metrics.

7.2.2.3 Per-individual quality control

All samples were screened to see if they had any non-synonymous variants in *SLC2A1* (this was in addition to the Custom Amplicon sequencing of *SLC2A1*).

Individuals with <20,000 exonic variants detected were removed (this would be <24,000 for Africans but no individuals with full African heritage were exome sequenced). With this criterion, most of the exome should be covered; from whole exome sequencing data, approximately 10,000-12,500 synonymous variants, 9,500-12,000 non-synonymous variants and 100–200 stop or splice altering variants are expected per individual, giving a total of approximately 24,000 SNVs in African American samples and 20,000 in European American samples [935, 945].

Individuals were not excluded based on mean coverage depth, a metric commonly used with whole genome sequencing data. This metric has almost no value in capture experiments, because the distribution of reads is uneven as a result of the capture [1045]. One threshold that has been previously adopted is $\geq 80\%$ bases with $\geq 20\times$ coverage [945, 1046], although, due to the variation in efficiency of exome enrichment, each exome must be sequenced to an average depth of 60-80 \times in order to obtain this level of coverage [945]. Instead, the percentage coverage of exonic bases at 2x/10x/20x/30x was inspected for each individual to detect outliers (by visual inspection of scatter plots created in Microsoft Excel 2010 (v. 14, Microsoft, Washington, USA)): see Figure 7.1, Figure 7.2, Figure 7.3 and Figure 7.4. Outliers are highlighted in circles; these individuals were removed from association analyses.

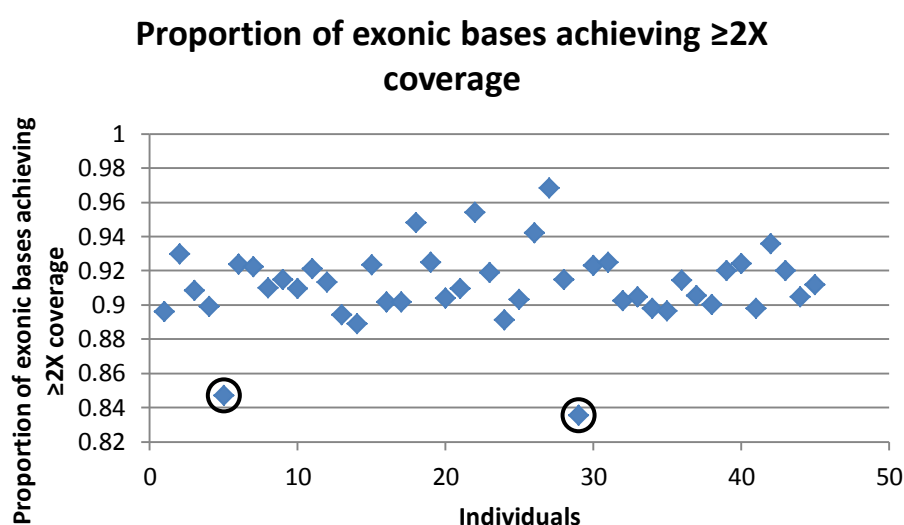


Figure 7.1: Proportion of exonic bases achieving $\geq 2\times$ coverage

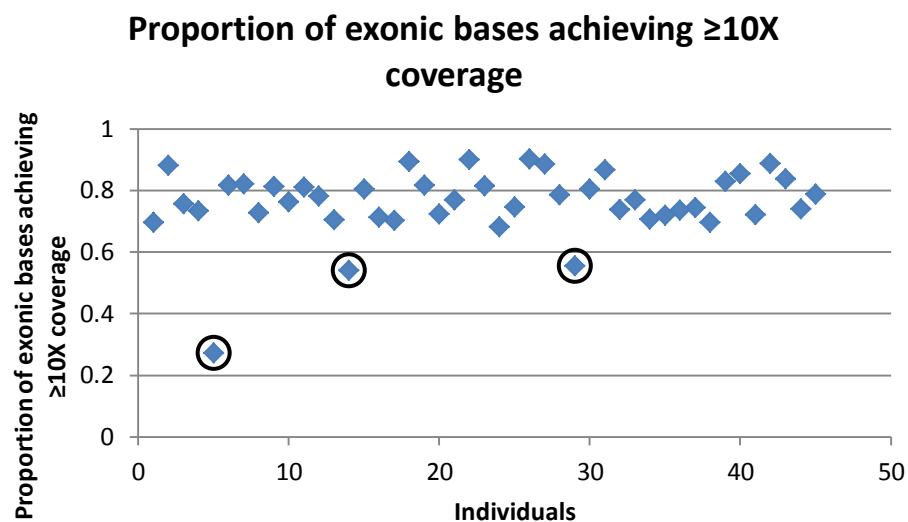


Figure 7.2: Proportion of exonic bases achieving $\geq 10X$ coverage

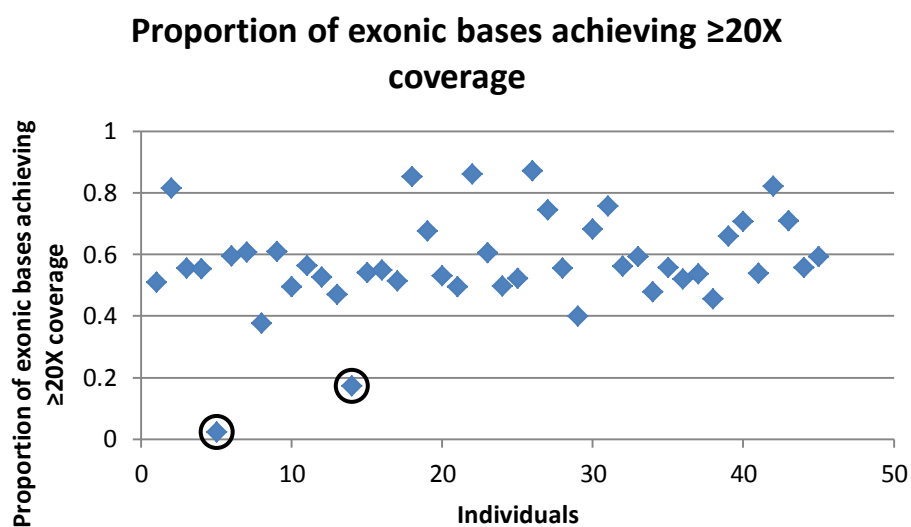


Figure 7.3: Proportion of exonic bases achieving $\geq 20X$ coverage

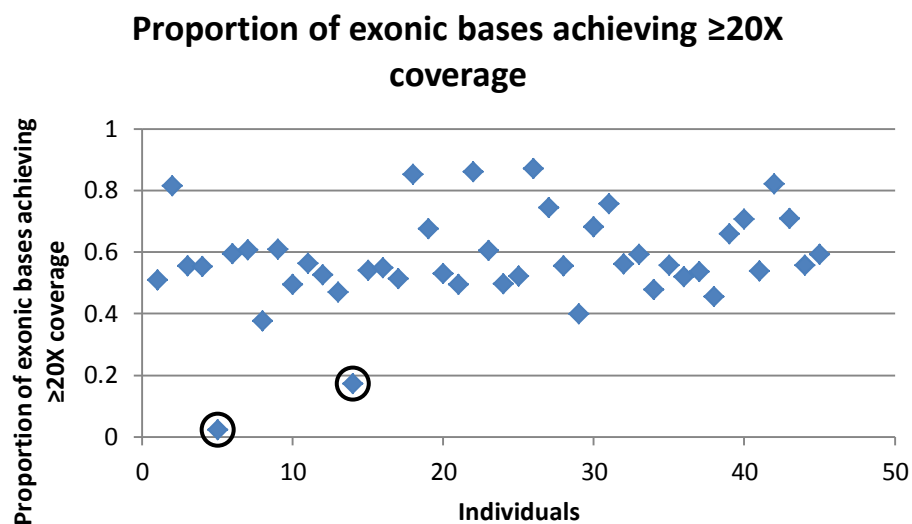


Figure 7.4: Proportion of exonic bases achieving $\geq 30X$ coverage

The mean Ti/Tv ratio, mean genotype quality score, the number of singletons and the proportion of heterozygous genotypes per individual were also plotted in a scatter graph to identify outliers due to potential sequencing errors (see Figure 7.5, Figure 7.6, Figure 7.7 and Figure 7.8).

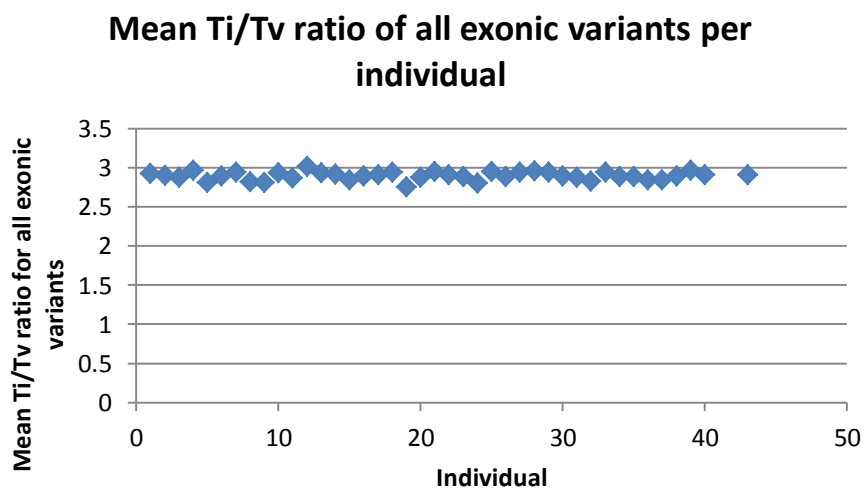


Figure 7.5: Ti/Tv ratio of all exonic variants per individual

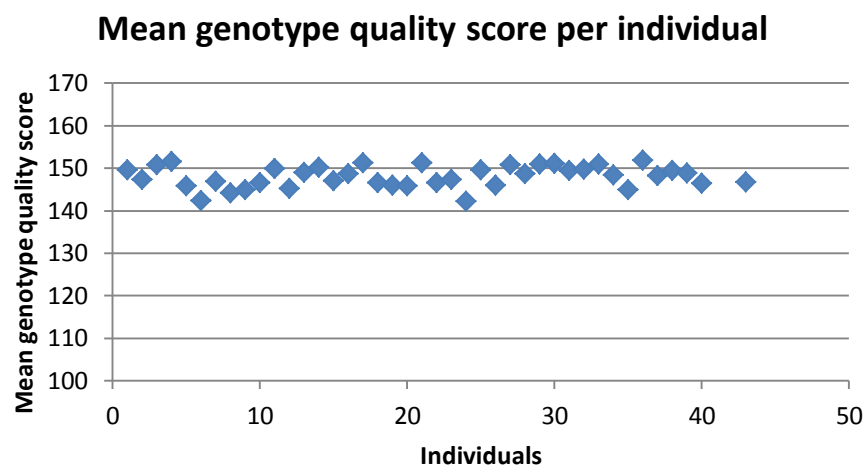


Figure 7.6: Mean genotype quality score per individual

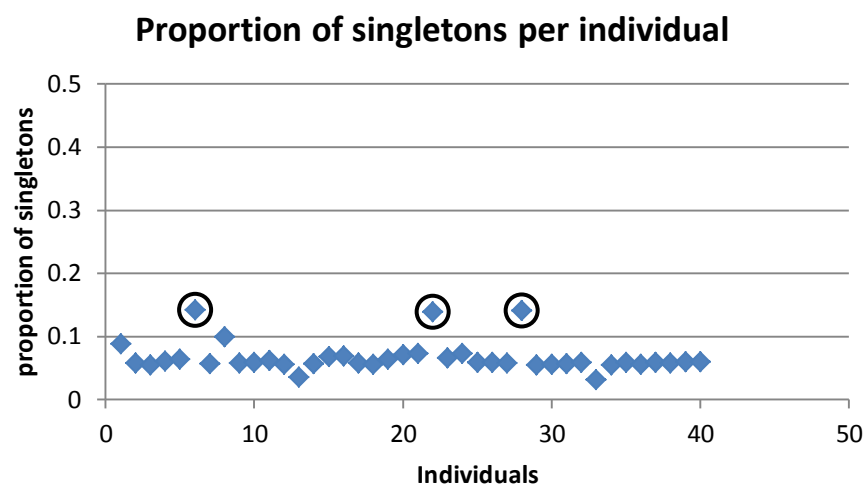


Figure 7.7: Proportion of singletons per individual

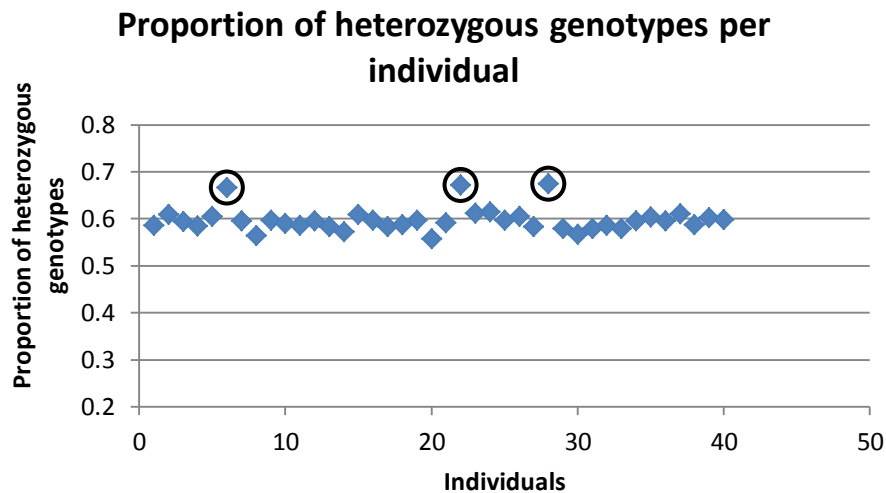


Figure 7.8: Proportion of heterozygous genotypes per individual

In order to exclude individuals with a greater degree of relatedness compared to others, for each pair of individuals who passed quality-control criteria, the number of genotypes for which both individuals carry an alternative allele was calculated in PLINK/SEQ (command: *ibs-matrix --long-format --two-counts*). The number of genotypes where both members of the pair carry an alternative allele divided by the total number of non-missing genotypes for the pair was plotted in a scatter graph. Outliers were determined by visual inspection of the plot (see Figure 7.9) and the individual from the pair with the lowest coverage was removed from association analyses.

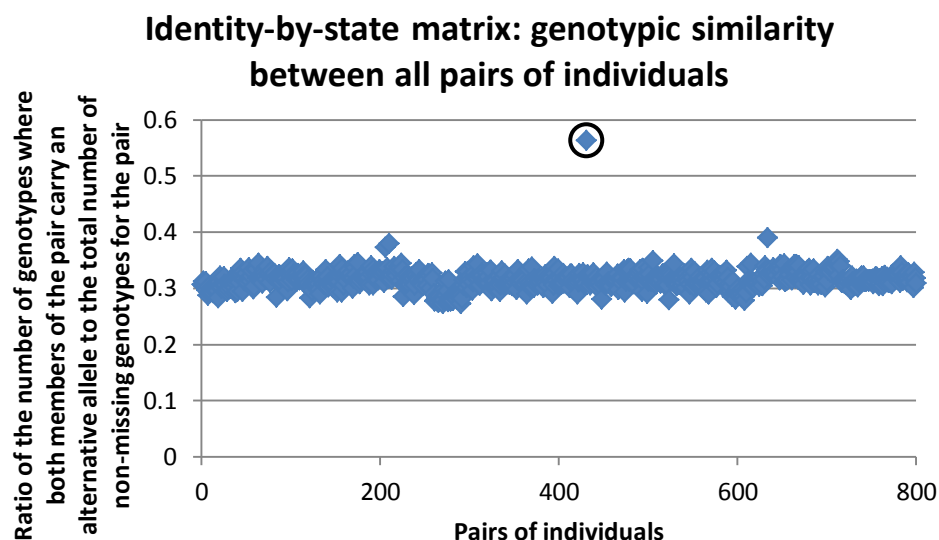


Figure 7.9: Identity-by-state matrix: genotypic similarity between all pairs of individuals

7.2.3 Population structure

In order not to bias association study results, differences in population structure must be balanced between responders and non-responders. There are several measures that can be used to check ancestry mismatching. Non-synonymous variants show greater levels of population differentiation than synonymous variants [1047] and the ratio of non-synonymous to synonymous variation varies between regions with different ancestries [814]. The proportion of heterozygous genotypes and novel variants also varies between populations: individuals of African ancestry show the highest rates of heterozygosity and novelty, compared to those of Native American and European ancestry [814]. The proportion of non-synonymous, non-synonymous:synonymous, novel and heterozygous variation may thus be used as a rough indication of ancestry. An unpaired t-test of unequal variance was used to determine whether the mean proportions were different between responders and non-responders. As outlined in Table 7.2, these metrics are not significantly different.

Table 7.2: Mean performance metrics for responders and non-responders as measures of ancestry matching

	Responders	Non-responders	P-value from t-test*
Proportion of novel exonic variants**	2.50	2.51	0.46
Proportion of non-synonymous variants	47.11	46.92	0.08
Ratio of non-synonymous:synonymous variants	1.05	1.06	0.10
Proportion of heterozygous genotypes	0.59	0.59	0.17

*unpaired Student's t-test with a two-tailed distribution calculated in Microsoft Excel

**not reported in dbSNP137, 1000 Genomes Project (April 2012 release), or NHLBI (data release ESP6500, with 6,503 individuals), according to ANNOVAR annotation

7.2.4 Association analyses

7.2.4.1 Single variant approach

A Fisher's exact test, excluding variants with deviation from HWE ($p=1 \times 10^{-6}$) was conducted in PLINK/SEQ (command: *v-assoc --mask hwe=0:1e-6*). In view of the small sample size, reference/alternative allele counts in responders and non-responders were examined, and p-values obtained from the association test.

λ_{GC} cannot be calculated in PLINK/SEQ. The vcf file was thus converted to PLINK format using Vcftools and the Fisher's exact test run in PLINK with the command *--fisher --adjust*.

7.2.4.2 Gene-based approach

As a complex trait, it may be that multiple variants in a gene (present in different people) influence KD response. A gene-based test is therefore appropriate, and it should give higher power to detect association signals from variants (especially those of low MAF or with small effect sizes), compared to single-variant tests [797] and there is evidence to suggest that rare variants often act collectively to influence disease risk [857, 1048, 1049].

Binary files were imported into PLINK/SEQ (v0.08, <http://atgu.mgh.harvard.edu/plinkseq/index.shtml>), where gene-based association tests were conducted (command: *assoc --mask loc.group=refGene --mac=1-9999*).

refGene refers to a reference gene region file that was uploaded to the locus database. This file, based on the RefSeq file provided by PLINK/SEQ (NCBI Reference Sequence Database), using hg19 coordinates, was checked and amended by Dr Costin Leu. `--mac=1-9999` specifies that only variants with between 1 and 9999 (any high number could be used in this case) copies of the alternative allele are included. P-values were obtained by adaptive permutation (permutation in PLINK/SEQ gives empirical p-values and does not account for multiple testing).

The c-alpha test [660] was first conducted (command: `assoc --tests calpha`), as it accounts for the possibility that a gene may contain both protective and deleterious (with relation to the phenotypic trait being analysed) variants and thus has higher power than one-directional tests when the association effects of variants in a gene are not uni-directional [660, 1036, 1050]. In this study, controls are not representative of the general population (the phenotypic trait is not a disease, it is response to a treatment) and so it cannot be assumed that susceptibility variants will be damaging. Mutations in *SLC2A1*, for example, are protective with regards KD response. Instead of using a burden-based summary statistic over a gene region, the c-alpha test compares the observed and expected distribution of minor alleles in cases and controls at each locus, assuming independence of each variant, to determine whether there is evidence of an 'unusual distribution'. Evidence from multiple loci is combined to form the test statistic. Singleton variants are pooled into a single binomial count per phenotype, akin to a burden test.

A gene burden and count of case-unique alleles test were also performed in PLINK/SEQ, as c-alpha was found to have low power in certain circumstances. These tests assume that all variants within a gene act in the same direction and that a higher burden of variants is damaging, rather than protective with regards the phenotypic trait. In the gene burden test, the number of non-reference alleles in a gene present in cases is compared to those in

controls; in the unique test, the number of case-unique non-reference alleles is compared. These tests were conducted once with non-responders as cases and once with responders as cases in order to identify genes that are associated with unfavourable and favourable KD response.

As well as examining p-values from gene-based tests, the 'DESC' columns in the output files were also examined to identify genes with the greatest differential distribution of alternative alleles between the responder and non-responder groups. In the c-alpha test, for example, this column gives a frequency breakdown of the number of variants with alternative alleles in cases and controls. The rationale for this is that, even if the differential distribution of alternative alleles in one gene amongst cases and controls does not generate a low p-value (for example, due to the small sample size), if all or many responders harbour the alternative alleles of variants in a particular gene but no or very few non-responders do, this suggests that the gene (or the variants within that gene) have a strong effect on the phenotype. The greater the number of samples in one response group with variants in that particular gene, the stronger the effect size these variants should have on the phenotype. For example, a gene with variants in all responders (or all but one/two/three responders and so on) but no (or very few) non-responders would certainly be of interest. If variants in a particular gene that influenced KD response were fully penetrant, they would be expected to be present in all members of one response group, compared to no members of the other response group. This may not be the case with KD response, but it is a prudent first step in the rare variant analysis process. The genes with the greatest differential distribution of variants in responders/non-responders or those with the lowest p-values in the gene-based tests were looked up in online databases (<http://www.ncbi.nlm.nih.gov/gene> and <http://www.genecards.org>) to determine whether they are known to be associated with neurological/metabolic function.

One set of association analyses (referred to as group 0) was conducted with all exonic and splicing variants, with no filtering based on MAF or predicted functional consequence. This accounts for the fact that variants that affect KD response may not be subject to negative selection pressure in the general population – response to starvation or metabolic therapies that mimic starvation may not affect fitness, particularly in populations with less exposure to periods of famine throughout history. The group 0 analyses also account for the fact that synonymous variants may also be of importance, although they have been less frequently associated with disease compared to non-synonymous variants.

Further analyses were then conducted with various filtering criteria, to include variants that are most likely to be functional:

- i) group 1: only including all common and rare functional exonic variants (non-synonymous [SNVs or indels in multiples of three], stop-gain, stop-loss, frameshift indels and splice site variants).
- ii) group 2: only including loss-of-function and splicing variants (stop-gain SNVs, frameshift indels and splice site variants). Loss-of-function SNVs are enriched for low-frequency alleles compared to synonymous and missense SNVs [815] and so this filtering approximates a rare variant analysis.

Variants were not filtered according to predictions of functional consequence by algorithms because of the uncertainty associated with such algorithms. They may be used as a guide to estimate the effects of the most interesting variants.

The novelty approach (where only previously-unidentified variants are included, a technique commonly used for Mendelian disease gene identification) was not used, as it assumes that the susceptibility variant is rare and likely to be previously unidentified.

7.2.4.3 Correction for multiple testing

Adjustment for multiple testing was made with a Bonferroni correction for the number of genes considered in each test: 0.05 divided by the number of genes that could generate a $p\text{-value} < 1$ (*i-value* [an estimate of the minimal achievable p -value for a gene given in PlinkSeq] < 1). This approach has been used in recent studies to set significance thresholds for gene-burden tests when assessing rare variant genotype-phenotype associations [714, 797, 944].

A suggestive significance threshold was set by dividing 1 by the number of tests, a strategy used in previous publications [693, 694].

7.2.4.4 Pathway-based analysis

As a complex trait, response to the KD may be influenced by many different genes, rather than just one single gene. A self-contained pathway analysis was conducted, testing the null hypothesis that no particular pathway is associated with KD response. A list of genes and their associated pathways was downloaded from the Broad Institute website (www.broadinstitute.org/gsea). This list included KEGG, PANTHER, REACTOME, BIOCARTA and PID pathways. This avoids potential bias from limiting oneself to one set of pathways. Analyses were not conducted with GO terms due to the general nature of many of these categories, which makes interpretation difficult.

The list of genes and associated pathways was uploaded into the locus database in PLINK/SEQ. Association analyses were performed with the same method used for gene-based association tests, but subsequently specifying the pathway with which the gene is associated (command: `assoc --mask locset.group=refGene,allPathways --mac=1-9999`). In this way, variants are grouped into genes (identified in the *refGene* file) and these genes are then grouped into pathways (located in the *allPathways* file). The gene burden and case-unique tests were conducted as the primary analyses (once with non-responders as

cases and once with responders as cases). The c-alpha test was not used in order to distinguish between those pathways that adversely affect KD response and those that favourably affect KD response.

7.3 Results

7.3.1 Quality control and population structure

One extreme responder was excluded due to presence of a non-synonymous variant in *SLC2A1*. This was subsequently confirmed in a clinical lab at Birmingham Children's Hospital.

Three individuals were excluded as <20,000 exonic variants were discovered and the proportion of variants with 2x/10x/20x/30x coverage was lower than in other samples. Another individual was excluded due to low coverage.

One pair of individuals were identified as outliers in terms of genotypic similarity. The individual with the lower coverage (2x, 10x, 20x and 30x) was removed.

Three individuals were excluded due to high proportions of heterozygous variants and number of singletons.

No individuals were excluded on the basis of Ti/Tv ratio or genotype quality.

No individuals were excluded based on proportions of novel exonic variants as, for all individuals, ≥96% of all called exonic variants were previously reported in public databases.

14 responders and 22 non-responders were included in the analyses.

Table 7.3: Function of variants included in the exome sequencing analyses

Variant function	Exonic function	Number of variants
Downstream	n/a	60
Exonic	Frameshift substitutions	256
	Non-frameshift substitutions	509
	Non-synonymous	39017
	Stop-gain	539
	Stop-loss	37
	Synonymous	35301
	Unknown	698
Exonic splicing	Frameshift substitutions	7
	Non-frameshift substitutions	4
	Non-synonymous	280
	Stop-gain	5
	Synonymous	211
	Unknown	9
Intergenic	n/a	2491
Intronic	n/a	1988
Exonic non-coding RNA	n/a	727
Intronic non-coding RNA	n/a	119
Splicing non-coding RNA	n/a	8
UTR3 non-coding RNA	n/a	7
UTR5 non-coding RNA	n/a	2
Splicing	n/a	318
Upstream	n/a	76
Upstream; downstream	n/a	4
UTR3	n/a	495
UTR5	n/a	253
Total		83421

Table 7.4: Performance metrics for all exonic variants for all samples included in analyses, n=36

Sample	Number of called exonic variants	Percentage of novel exonic variants*	Percentage of non-synonymous variants	Mean Ti/Tv ratio	Percentage of heterozygous genotypes	Percentage of singletons	Mean quality score
1	20717	4.04	46.99	2.92	58.55	8.72	149.53
2	22207	2.34	47.24	2.90	60.91	5.72	147.33
3	20852	2.56	46.81	2.87	59.40	5.44	150.83
4	20290	2.24	46.66	2.97	58.46	6.03	151.57
5	21563	2.62	46.61	2.81	60.54	6.29	145.78
6	21267	2.09	46.80	2.94	59.58	5.62	146.89
7	21134	3.73	46.98	2.82	56.39	9.90	144.18
8	21534	3.66	46.95	2.81	59.67	5.76	144.85
9	21235	2.20	46.80	2.93	58.91	5.77	146.58
10	20159	2.52	47.18	2.87	58.52	6.09	149.94
11	21184	2.26	47.23	3.02	59.49	5.48	145.24
12	20489	2.39	47.37	2.93	58.25	3.44	148.96
13	20292	2.41	46.71	2.92	57.18	5.66	150.19
14	22713	2.48	47.08	2.84	60.87	6.77	147.09
15	21686	3.02	47.57	2.89	59.66	6.80	148.65
16	20413	2.38	46.74	2.91	58.35	5.76	151.23
17	20410	1.97	46.94	2.94	58.65	5.52	146.59

18	22748	2.44	47.58	2.75	59.69	6.30	145.91
19	21560	3.61	47.27	2.87	55.71	7.06	145.77
20	20659	2.80	46.76	2.95	59.17	7.21	151.27
21	22779	2.48	46.89	2.89	61.19	6.50	147.37
22	22758	2.50	48.02	2.81	61.44	7.21	142.28
23	21649	2.38	46.80	2.95	59.71	5.82	149.56
24	22113	2.35	47.20	2.89	60.51	5.81	145.93
25	20531	2.25	47.00	2.94	58.24	5.75	150.84
26	20151	2.00	47.04	2.94	57.95	5.39	150.91
27	20547	2.56	46.52	2.89	56.69	5.54	151.05
28	20472	2.12	46.93	2.88	57.85	5.57	149.37
29	20802	2.45	46.71	2.83	58.53	5.85	149.74
30	21031	2.17	46.91	2.89	59.59	5.45	148.40
31	22019	2.39	47.09	2.89	60.33	5.80	144.10
32	20171	2.43	46.38	2.84	59.54	5.51	151.90
33	21656	2.36	47.33	2.85	61.06	5.78	148.21
34	20601	2.33	46.56	2.89	58.70	5.69	149.44
35	21018	2.47	47.21	2.97	60.19	5.91	148.79
36	21951	2.29	46.92	2.91	59.81	5.97	146.46

*not reported in dbSNP137, 1000 Genomes Project (April 2012 release), or NHLBI (data release ESP6500, with 6,503 individuals), according to ANNOVAR annotation

7.3.2 Association studies

7.3.2.1 Single variant approach

λ_{GC} was 1.00, indicative of lack of importance of population structure.

A significance threshold of 6.42×10^{-7} was set, based on testing for 84905 variants. No statistically significant results were obtained, nor were any p-values below the suggestive significance threshold of 1.177×10^{-5} (1/number of variants or tests). The lowest p-value was 0.00019998, from rs3986599, a non-synonymous SNV in *TRPM5*, which was present in 17/22(77%) non-responders and 1/14 responders, all with genotype A/G.

Other variants with the largest differences in alternative allele count between responders and non-responders were:

rs5005869, a non-synonymous SNV in *ANKRD36C*, present in 17/22(77%) non-responders and 1/14(7%) responders, all with genotype C/T.

rs79074863, a non-synonymous SNV in *XPO1*, present in 13/22(59%) non-responders and 0 responders.

rs2490085, a SNV in the 5'UTR of *CCNYL2*, present in 12/22(%) non-responders and 0 responders, all with genotype A/T.

rs77242743, a synonymous SNV in *BMP1*, present in 12/22(%) non-responders and 0 responders, all with genotype A/C.

NIPA2, *L2HGDH*, *ADM5*, *NBPF8* and *CDH23* each had a variant present in 11/22(50%) non-responders and 0 responders. Only those variants present in *L2HGDH* (c.T593G:p.V198G at chr14:50750699; all individuals had genotype A/C) and *ADM5* (c.G344C:p.R115P at chr19:50193632, all individuals had genotype C/G) were exonic.

rs79074863 and c.T593G:p.V198G are likely to affect gene function. They are predicted to be damaging by all functional algorithms in ANNOVAR (SIFT score=0, PolyPhen2 score=1, LRT score=1, MutationTaster score=1) and are located in conserved regions (GERP++ scores=5.43 and 5.22 and PhyloP scores=0.999). The effects of the other variants on gene function are unknown: rs3986599 is not predicted to be damaging by any of the algorithms in ANNOVAR and it is not highly conserved (GERP++ score=1.57); rs5005869 has a SIFT score of 0.35 (the closer the score is to 0, the more damaging the variant) but no other ANNOVAR annotation; no annotation was available for rs2490085 or c.G344C:p.R115P.

None of these variants have been associated with any disease/traits, according to Phenotype Genotype Integrator.

Of the genes including coding variants, *L2HGDH*, *TRPM5* and *XPO1* have a role that could feasibly influence KD response.

L2HGDH encodes L-2-hydroxyglutarate dehydrogenase, an enzyme that oxidizes L-2-hydroxyglutarate to alpha-ketoglutarate, which is an intermediate in the Krebs cycle. Mutations in *L2HGDH* cause the neurometabolic disorder L-2-hydroxyglutaric aciduria.

TRPM5 encodes a calcium-activated cation channel (transient receptor potential cation channel subfamily M member 5), which plays a role in the signal transduction for bitter and sweet taste in sensory neurons [1051] and has been shown to regulate glucose-stimulated insulin secretion: *Trpm5*^{-/-} mice had prolonged elevation of blood glucose levels compared to wild-type mice after a glucose load [1052].

XPO1 encodes exportin 1, which mediates nuclear export signal-dependent protein transport. This protein has been shown to mediate cellular response to stress and may influence glucose homeostasis.

Little is known about the role of *ANKRD36C*. *ADM5* is thought to encode a non-functional remnant of adrenomedullin-5, which is a short chain of amino acids associated with tumours of the medulla of the adrenal glands.

7.3.2.2 Gene-based approach

C-alpha analyses:

- i) Group 0: all exonic and splicing variants, no filtering.
- ii) Group 1: missense exonic and splicing variants - non-synonymous (SNV or indels in multiples of 3), stop-gain, stop-loss, frameshift indels and splice site variants.
- iii) Group 2: exonic loss-of-function and splicing variants.

Group 0: 16314 genes with i-value <1 = significance threshold of 3.06×10^{-6} and a suggestive significance threshold of 6.13×10^{-5} . *ANKRD36C* reached statistical significance ($p=2.00 \times 10^{-6}$) and five genes reached suggestive significance.

Group 1: 13167 genes with i-value <1 = significance threshold of 3.79×10^{-6} and a suggestive significance threshold of 7.59×10^{-5} . Seven genes reached suggestive significance.

Group 2: 1662 genes with i-value <1 = significance threshold of 3.008×10^{-5} and a suggestive significance threshold of 0.000601. No genes reached significance or suggestive significance.

Genes that reached significance or suggestive significance in c-alpha Group 0 and 1 analyses are given in Table 7.5.

Table 7.5: Genes that reached significance or suggestive significance in c-alpha analyses: Groups 0 and 1

Group 0				
Gene	Number of variants in gene	P-value	I-value	Frequency breakdown: non-responder/responder alternative allele counts for particular variants (variant class and rs number or position given from ANNOVAR annotation)
<i>ANKRD36C</i>	48	2.00x10 ⁻⁶	1.00x10 ⁻⁶	<p>8 variants present in 0 non-responders and 1 responder: synonymous SNV (rs201755982), non-synonymous SNV at chr2:96521262, non-synonymous SNV (rs74328999), synonymous SNV (rs10211258), non-synonymous SNV (rs78780931), synonymous SNV at chr2:96591119, frameshift insertion at chr2:96610401, non-synonymous SNV at chr2:96557418</p> <p>3 variants present in 0 non-responders and 2 responders: stop-gain SNV (rs74941794), synonymous SNV (rs183934154), splicing variant (rs202102082)</p> <p>1 variant present in 0 non-responders and 5 responders: non-synonymous SNV rs80350011</p> <p>9 variants present in 1 non-responder and 0 responders: non-synonymous SNV (rs200478637), non-synonymous SNV (rs5005871), non-synonymous SNV (rs2951081), non-synonymous SNV (rs112858216), frameshift deletion at chr2:96525668..96525672, synonymous SNV (rs189877528) non-synonymous SNV (rs199969223), synonymous SNV at chr2:96601236), synonymous SNV (rs115480888)</p> <p>1 variant present in 1 non-responder and 1 responder: non-synonymous SNV (rs78585559)</p> <p>1 variant present in 1 non-responder and 2 responders: synonymous SNV (rs75657861)</p> <p>1 variant present in 10 non-responders and 2 responders: non-synonymous SNV (rs78473403)</p> <p>1 variant present in 12 non-responders and 2 responders: splicing variant (rs1819101)</p> <p>1 variant present in 14 non-responders and 13 responders: non-synonymous SNV (rs62154594)</p> <p>1 variant present in 16 non-responders and 9 responders: non-synonymous SNV (rs74946480)</p> <p>1 variant present in 17 non-responders and 1 responder: rs5005869 non-synonymous SNV)*</p> <p>1 variant present in 17 non-responders and 13 responders: non-synonymous SNV (rs78739183)</p> <p>1 variant present in 17 non-responders and 14 responders: non-synonymous SNV (rs77710257)</p> <p>1 variant present in 17 non-responders and 9 responders: synonymous SNV (rs1636081)</p> <p>1 variant present in 18 non-responders and 12 responders: non-synonymous SNV (rs62154562)</p> <p>3 variants present in 19 non-responders and 14 responders: synonymous SNV (rs62154593), non-synonymous SNV (rs62154561), non-synonymous SNV (rs62154618)</p> <p>1 variant present in 2 non-responders and 2 responders: synonymous SNV (rs202113195)</p> <p>1 variant present in 2 non-responders and 4 responders: non-synonymous SNV (rs35461149)</p> <p>1 variant present in 20 non-responders and 12 responders: non-synonymous SNV (rs1636083)</p> <p>1 variant present in 21 non-responders and 13 responders: non-synonymous SNV (rs199988716)</p>

				1 variant present in 21 non-responders and 8 responders: synonymous SNV (rs5005870) 1 variant present in 22 non-responders and 14 responders: synonymous SNV (rs148485539) 1 variant present in 3 non-responders and 2 responders: synonymous SNV (rs4907220) 1 variant present in 3 non-responders and 5 responders: non-synonymous SNV (rs79908880) 1 variant present in 4 non-responders and 7 responders: synonymous SNV (rs10211258) 1 variant present in 5 non-responders and 0 responders: synonymous SNV (rs62156871) 1 variant present in 6 non-responders and 0 responders: non-synonymous SNV (rs5005868) 1 variant present in 8 non-responders and 7 responders: synonymous SNV (rs10211258) 1 variant present in 9 non-responders and 6 responders: non-synonymous SNV (rs11695514)
<i>CCDC104</i>	3	2.70×10^{-5}	3.85×10^{-6}	1 variant present in 1 non-responder and 10 responders: non-synonymous SNV (rs3762513) 1 variant present in 3 non-responders and 1 responder: non-synonymous SNV (rs1045920) 1 variant present in 3 non-responders and 2 responders: non-synonymous SNV (rs1045910)
<i>TBC1D3</i>	6	3.14×10^{-5}	2.24×10^{-6}	3 variants present in 0 non-responders and 1 responder: non-synonymous SNV (rs11550750), non-synonymous (rs199547231), synonymous SNV (rs4606788) 1 variant present in 0 non-responders and 5 responders: non-synonymous SNV (rs4795237) 1 variant present in 5 non-responders and 12 responders: non-synonymous SNV (rs306818) 1 variant present in 6 non-responders and 0 responders: synonymous SNV (rs306820)
<i>TRPM5</i>	11	3.22×10^{-5}	2.30×10^{-6}	1 variant present in 0 non-responders and 1 responder: frameshift insertion at chr11:2434061 1 variant present in 1 non-responder and 0 responders: synonymous SNV (rs141187622) 1 variant present in 1 non-responder and 2 responders: non-synonymous SNV (rs34364959) 1 variant present in 10 non-responders and 0 responders: synonymous SNV (rs4929981) 1 variant present in 10 non-responders and 3 responders: non-synonymous SNV (rs4929982) 1 variant present in 12 non-responders and 9 responders: synonymous SNV (rs2074234) 1 variant present in 17 non-responders and 1 responder: non-synonymous SNV (rs3986599) 1 variant present in 2 non-responders and 0 responders: non-synonymous SNV (rs34551253) 1 variant present in 2 non-responders and 2 responders: non-synonymous SNV (rs34350821) 1 variant present in 4 non-responders and 5 responders: synonymous SNV (rs800342) 1 variant present in 9 non-responders and 9 responders: non-synonymous SNV (rs886277)
<i>HMGXB4</i>	2	3.33×10^{-5}	2.38×10^{-6}	1 variant present in 1 non-responder and 0 responders: frameshift insertion at chr22:35661553 1 variant present in 2 non-responders and 11 responders: non-synonymous SNV (rs1053593)
<i>C1orf105</i>	10	4.20×10^{-5}	3.00×10^{-6}	1 variant present in 0 non-responders and 1 responder: non-synonymous SNV at chr1:172411009* 3 variants present in 1 non-responder and 0 responders: non-synonymous SNV at chr1:172410881*, non-synonymous SNV (rs139238043)*, synonymous SNV (rs2285175)* 1 variant present in 1 non-responder and 1 responder: non-synonymous SNV (rs34001453)* 2 variants present in 1 non-responder and 4 responders: splicing variant (rs41310899), non-synonymous SNV

				(rs16844498) 1 variant present in 1 non-responder and 9 responders: non-synonymous SNV (rs1129942) 1 variant present in 7 non-responders and 12 responders: non-synonymous SNV (rs1063412)* 1 variant present in 7 non-responders and 7 responders: synonymous SNV (rs2230471)*
Group 1				
Gene	Number of variants in gene	P-value	I-value	Frequency breakdown: non-responder/responder alternative allele counts for particular variants (variant class and rs number or position given from ANNOVAR annotation)
ANKRD36C	33	1.20x10 ⁻⁵	1.00x10 ⁻⁶	5 variants present in 0 non-responders and 1 responder: non-synonymous SNV at chr2:96521262, non-synonymous SNV (rs74328999), non-synonymous SNV (rs78780931), frameshift insertion at chr2:96610401, non-synonymous SNV at chr2:96557418 2 variants present in 0 non-responders and 2 responders: stop-gain SNV (rs74941794), splicing variant (rs202102082) 1 variant present in 0 non-responders and 5 responders: non-synonymous SNV rs80350011 9 variants present in 1 non-responder and 0 responders: non-synonymous SNV (rs200478637), non-synonymous SNV (rs5005871), non-synonymous SNV (rs2951081), non-synonymous SNV (rs112858216), frameshift deletion at chr2:96525668..96525672, non-synonymous SNV (rs199969223) 1 variant present in 1 non-responder and 1 responder: non-synonymous SNV (rs78585559) 1 variant present in 10 non-responders and 2 responders: non-synonymous SNV (rs78473403) 1 variant present in 12 non-responders and 2 responders: splicing variant (rs1819101) 1 variant present in 14 non-responders and 13 responders: non-synonymous SNV (rs62154594) 1 variant present in 16 non-responders and 9 responders: non-synonymous SNV (rs74946480) 1 variant present in 17 non-responders and 1 responder: non-synonymous SNV (rs5005869) 1 variant present in 17 non-responders and 13 responders: non-synonymous SNV (rs78739183) 1 variant present in 17 non-responders and 14 responders: non-synonymous SNV (rs77710257) 1 variant present in 18 non-responders and 12 responders: non-synonymous SNV (rs62154562) 2 variants present in 19 non-responders and 14 responders: non-synonymous SNV (rs62154561), non-synonymous SNV (rs62154618) 1 variant present in 2 non-responders and 4 responders: non-synonymous SNV (rs35461149) 1 variant present in 20 non-responders and 12 responders: non-synonymous SNV (rs1636083) 1 variant present in 21 non-responders and 13 responders: non-synonymous SNV (rs199988716)3/5(1) 1 variant present in 6 non-responders and 0 responders: non-synonymous SNV (rs5005868) 1 variant present in 8 non-responders and 7 responders: synonymous SNV (rs10211258) 1 variant present in 9 non-responders and 6 responders: non-synonymous SNV (rs11695514)

<i>HMGXB4</i>	2	1.56x10 ⁻⁵	2.22x10 ⁻⁶	1 variant present in 1 non-responder and 0 responders: frameshift insertion at chr22:35661553 1 variant present in 2 non-responders and 11 responders: non-synonymous SNV (rs1053593)
<i>C1orf105</i>	8	1.56x10 ⁻⁵	1.11x10 ⁻⁶	1 variant present in 0 non-responders and 1 responder: non-synonymous SNV at chr1:172411009* 2 variants present in 1 non-responder and 0 responders: non-synonymous SNV at chr1:172410881*, synonymous SNV (rs2285175)* 1 variant present in 1 non-responder and 1 responder: non-synonymous SNV (rs34001453)* 2 variants present in 1 non-responder and 4 responders: splicing variant (rs41310899), non-synonymous SNV (rs16844498) 1 variant present in 1 non-responder and 9 responders: non-synonymous SNV (rs1129942) 1 variant present in 7 non-responders and 12 responders: non-synonymous SNV (rs1063412)*
<i>CCDC104</i>	3	3.33x10 ⁻⁵	2.38x10 ⁻⁶	1 variant present in 1 non-responder and 10 responders: non-synonymous SNV (rs3762513) 1 variant present in 3 non-responders and 1 responder: non-synonymous SNV (rs1045920) 1 variant present in 3 non-responders and 2 responders: non-synonymous SNV (rs1045910)
<i>TBC1D3</i>	4	3.35x10 ⁻⁵	4.78x10 ⁻⁶	2 variants present in 0 non-responders and 1 responder: non-synonymous SNV (rs11550750), non- synonymous (rs199547231) 1 variant present in 0 non-responders and 5 responders: non-synonymous SNV (rs4795237) 1 variant present in 5 non-responders and 12 responders: non-synonymous SNV (rs306818)
<i>IFI27L1</i>	3	4.49x10 ⁻⁵	9.94x10 ⁻⁵	1 variant present in 0 non-responders and 5 responders: non-synonymous SNV at chr14:94568370 1 variant present in 0 non-responders and 2 responders: non-synonymous SNV (rs117094493) 1 variant present in 0 non-responders and 5 responders: stop-gain SNV (rs79237142)
<i>TRPM5</i>	7	4.54x10 ⁻⁵	3.24x10 ⁻⁶	1 variant present in 0 non-responders and 1 responder: frameshift insertion at chr11:2434061 1 variant present in 1 non-responder and 2 responders: non-synonymous SNV (rs34364959) 1 variant present in 10 non-responders and 3 responders: non-synonymous SNV (rs4929982) 1 variant present in 12 non-responders and 9 responders: synonymous SNV (rs2074234) 1 variant present in 2 non-responders and 0 responders: non-synonymous SNV (rs34551253) 1 variant present in 2 non-responders and 2 responders: non-synonymous SNV (rs34350821) 1 variant present in 9 non-responders and 9 responders: non-synonymous SNV (rs886277)

*variants also overlap with *PIGC* gene

In this dataset, the c-alpha test was found to have low power to detect associations with genes containing variants that are all detrimental with regards KD response. With such genes, lower p-values are generated with tests that only allow for damaging variants. For example, when conducting the gene burden and count of case-unique alleles tests, which assume a uni-directional effect of variants, *XPO1* generated the lowest p-value:

Group 0

Burden test $p=3.59 \times 10^{-5}$

Count of case-unique alleles test $p=3.73 \times 10^{-5}$

compared to c-alpha test $p=9.30 \times 10^{-5}$ (just below suggestive significance)

XPO1 reached suggestive significance in the burden and count of case-unique alleles tests. rs143005485, a synonymous SNV, was present in two non-responders and 0 responders) and rs79074863, a non-synonymous SNV, was present in 13 non-responders and 0 responders (as identified from the single variant approach).

Group 1

XPO1 also generated the lowest p-value in the Group 1 burden and count of case-unique alleles tests, although it did not reach suggestive significance.

Group 2

No individual in this cohort harboured loss-of-function or splicing variants in *XPO1*. *CDC27* generated the lowest p-value in the burden test. *OR52J3* generated the lowest p-value in the count of case-unique alleles test. No genes reached significance or suggestive significance in any test.

Testing for protective variants (this assumes that harbouring the alternative allele is beneficial for KD response):

Group 0

POTEB ($p=3.59 \times 10^{-5}$) and *SGOL2* ($p=5.63 \times 10^{-5}$) reached suggestive significance in the burden test; *OR5M10* ($p=1.43 \times 10^{-5}$) and *IFI27L1* ($p=5.70 \times 10^{-5}$) reached suggestive significance in the case-unique alleles test.

Group 1

TBC1D3 ($p=2.08 \times 10^{-5}$) and *C1orf105* ($p=2.98 \times 10^{-5}$) reached suggestive significance in the burden test; *OR5M10* ($p=1.78 \times 10^{-5}$) and *IFI27L1* ($p=3.35 \times 10^{-5}$) reached suggestive significance in the case-unique alleles test.

Group 2

No genes reached significance or suggestive significance in any test. *PAPPA* had the lowest p-value in the burden test. *IFI27L1* had lowest p-value in the in the case-unique alleles test.

Examining alternative allele counts in responders/non-responders, *XPO1* is the gene in which the highest number of non-responders (and no responders) harbours variants. 12/22(55%) non-responders and no responders harbour variants in *ADM5*.

Other genes (not with case or control-unique variants) that may be of interest are *GBF1* and *APBA2*, in which 12/22(55%) non-responders and one responder harbour variants.

Information (taken from www.genecards.org, <http://www.ncbi.nlm.nih.gov> and literature searches) regarding the roles of genes that reached significance or suggestive significance or those genes with the greatest difference in the number of responders/non-responders harbouring alternative alleles is given in Table 7.6.

Table 7.6: Genes of most interest from exome sequencing analyses

Gene	Name	Role	GO terms	Pathways
<i>XPO1</i>	exportin 1 (CRM1 homolog, yeast)	Encodes exportin 1, which mediates leucine-rich nuclear export signal-dependent protein transport. The protein may influence glucose homeostasis via translocation of hepatic glucokinase [1053, 1054] and lead to drug-resistance in cancer [1055, 1056].	Receptor activity, intracellular protein transport, nucleobase, nucleoside, nucleotide and nucleic acid transport, transcription from RNA, polymerase II promoter, mRNA transcription, protein metabolic process	REACTOME pathways: cell cycle, mitotic; influenza reaction, HIV infection,
<i>ANKRD36C</i>	ankyrin repeat domain 36C	Unknown. Associated diseases are osteomalacia and uremia.	Protein binding, ion channel inhibitor activity	-
<i>TRPM5</i>	transient receptor potential cation channel subfamily M member 5	Encodes a member of the transient receptor potential protein family that an important role in taste transduction, and has characteristics of a calcium-activated, non-selective cation channel. It carries Na ⁺ , K ⁺ , and Cs ⁺ ions equally well, but not Ca(2 ⁺) ions. TRPM5 has been shown to affect blood glucose levels [1052].	receptor activity ion channel activity	KEGG pathways: Taste transduction
<i>CCDC104</i>	coiled-coil domain containing 104	May act as an effector for ADP-ribosylation factor-like 3 (ARL3) [1057]. It has been shown to increase susceptibility to <i>Citrobacter</i> infection [1058].	Protein binding	-
<i>TBC1D3</i>	TBC1 domain family, member 3	Encodes an oncogene that is over-represented in many different types of cancer, and controls cell proliferation [1059]. Acts as a GTPase activating protein for RAB5. It delays IRS1 degradation and increases insulin/IGF1-induced signal transduction [1060].	hydrolase activity, protein binding, small GTPase regulator activity, intracellular protein transport, exocytosis, cellular component morphogenesis	-
<i>HMGXB4</i>	HMG box domain containing 4	Encodes a nonhistone chromosomal protein that is thought to negatively regulate Wnt/beta-catenin signalling during development	transcription factor activity, chromatin binding, receptor binding, intracellular signalling cascade, regulation of transcription from RNA	PANTHER pathways: p53 pathway-High mobility group protein 1

			polymerase II promoter, establishment or maintenance of chromatin architecture	
<i>C1orf105</i>	chromosome 1 open reading frame 105	Uncharacterised	-	-
<i>CCNYL2</i>	cyclin Y-like 2	A pseudogene.	Protein kinase binding, regulation of cyclin-dependent protein serine/threonine kinase activity	-
<i>ADM5</i>	adrenomedullin 5 (putative)	Probable non-functional remnant of adrenomedullin-5	cysteine-type peptidase activity, ubiquitin-protein ligase activity, RNA splicing factor activity, transesterification mechanism mRNA binding gamete generation, nuclear mRNA splicing via spliceosome, proteolysis	-
<i>GBF1</i>	brefeldin A resistant guanine nucleotide exchange factor 1	Encodes a guanine nucleotide exchange factor, localised to the Golgi apparatus, that plays a role in vesicular trafficking by activating ADP ribosylation factor 1 (ARF5). ARF5 interacts with GLUT-4. Diseases associated with this gene include hepatitis c and Alzheimer's disease. Among its related super-pathways are 'Golgi Associated Vesicle Biogenesis and Cholesterol and Sphingolipids transport' and 'Transport from Golgi and ER to the apical membrane'.	protein binding, small GTPase regulator activity, guanyl-nucleotide exchange factor activity, amino acid transport, intracellular protein transport, exocytosis, cellular amino acid catabolic process	REACTOME pathways: Membrane trafficking
<i>APBA2</i>	beta (A4) precursor protein-binding, family A, member 2	Encodes a neuronal adapter protein that stabilises amyloid precursor protein (APP) and inhibits production of proteolytic APP fragments, including the A beta peptide that is deposited in the brains of Alzheimer's disease patients. The protein is also thought to be play a role in vesicular trafficking protein in the brain by binding to STXBP1.	protein binding and beta-amyloid binding, intracellular protein transport	PANTHER pathways: Alzheimer disease-amyloid secretase pathway-Amyloid beta A4 precursor protein-binding family A; Alzheimer disease-

				amyloid secretase pathway-HIV-1 Tat interactive protein, 60kDa
<i>POTEB</i>	POTE ankyrin domain family, member B	Associated with prostatitis and prostate cancer.	plasma membrane, extrinsic to membrane	-
<i>SGOL2</i>	shugoshin-like 2 (S. pombe)	Involved in chromosome segregation during mitosis and meiosis	M phase, cell cycle, chromosome segregation, sister chromatid cohesion, meiosis, protein localization, cell cycle process, cell cycle phase, meiotic chromosome segregation, meiotic sister chromatid cohesion, chromosome organization, cell division, meiotic cell cycle, M phase of meiotic cell cycle, meiotic sister chromatid cohesion, centromeric	REACTOME pathways: cell cycle, mitotic
<i>OR5M10</i>	olfactory receptor, family 5, subfamily M, member 10	Encodes an olfactory receptor.	cell surface receptor linked signal transduction, G-protein coupled receptor protein signaling pathway, sensory perception, sensory perception of chemical stimulus, sensory perception of smell, neurological system process, cognition, olfactory receptor activity	KEGG pathways: Olfactory transduction Reactome pathway: Signalling by GPCR
<i>IFI27L1</i>	interferon, alpha-inducible protein 27-like 1	Associated with hepatitis.	Integral component of membrane	

There are several predominant functions/themes amongst the genes in Table 7.6:

- i) Protein binding/transport (*TBC1D3*, *XPO1*, *ANKRD36C*, *TRPM5*, *CCDC104*, *CCNYL2*, *GBF1* and *APBA2*).
- ii) GTP-ase regulation (*CCDC104*, *GBF1* and *TBC1D3*)
- iii) Cell cycling (*SGOL2*, *CCNYL2* and *XPO1*)
- iv) Glucose transport/homeostasis and insulin signalling (*XPO1*, *TRPM5*, *GBF1* [activates ARF5, which interacts with GLUT-4], and *TBC1D3*).
- v) Neurological processes (*APBA2*, *GBF1* and *OR5M10*).

It is unknown what effects the variants in any of the genes of interest have on gene function. Only four variants were predicted to be damaging and conserved by all prediction algorithms given in ANNOVAR: rs79074863 in *XPO1*, c.T593G:p.V198G in *L2HGDH* (both as previously noted), rs1063412 and the non-synonymous SNV at chr1:172410881, both in *C1orf105* (intronic region) and *PIGC* (exonic region).

7.3.2.3 Pathway analyses

With a significance threshold of $0.05/1320=3.79 \times 10^{-5}$, no pathway was significantly associated with either favourable or unfavourable KD response with the burden test.

When testing for association with unfavourable KD response, the PID ‘ERBB1 Internalisation’ pathway reached significance ($p=2.64 \times 10^{-5}$) in the count of case-unique alleles test. The BIOCARTA ‘PTC1’ pathway generated the second lowest p-value, which reached suggestive significance (threshold= $1/1320=0.00076$, with a p-value of 4.67×10^{-5}).

When testing for association with favourable KD response, the REACTOME pathway ‘Triglyceride biosynthesis’ reached suggestive significance in the burden test.

The PID ‘ERBB1 Internalisation’ pathway is very large (a figure displaying the pathway can be found at <http://pid.nci.nih.gov>). This pathway includes 41 genes: *AMPH*, *ARHGEF7*, *CBL*,

CBLB, CDC42, CHMP3, DNM1, EGF, EGFR, EPN1, EPS15, GRB2, HGS, HRAS, ITSN1, KRAS, LRIG1, NRAS, PIK3CA, PIK3CB, PIK3CD, PIK3R1, PIK3R2, PIK3R3, PTK2, RAB5A, RAF1, SH3GL2, SH3KBP1, SHC1, SOS1, SPRY2, SRC, STAMBP, SYNJ1, TSG101, UBE2D1, UBE2D2, UBE2D3, USP8, and ZFYVE28.

The PTC1 pathway is outlined in Figure 7.10.

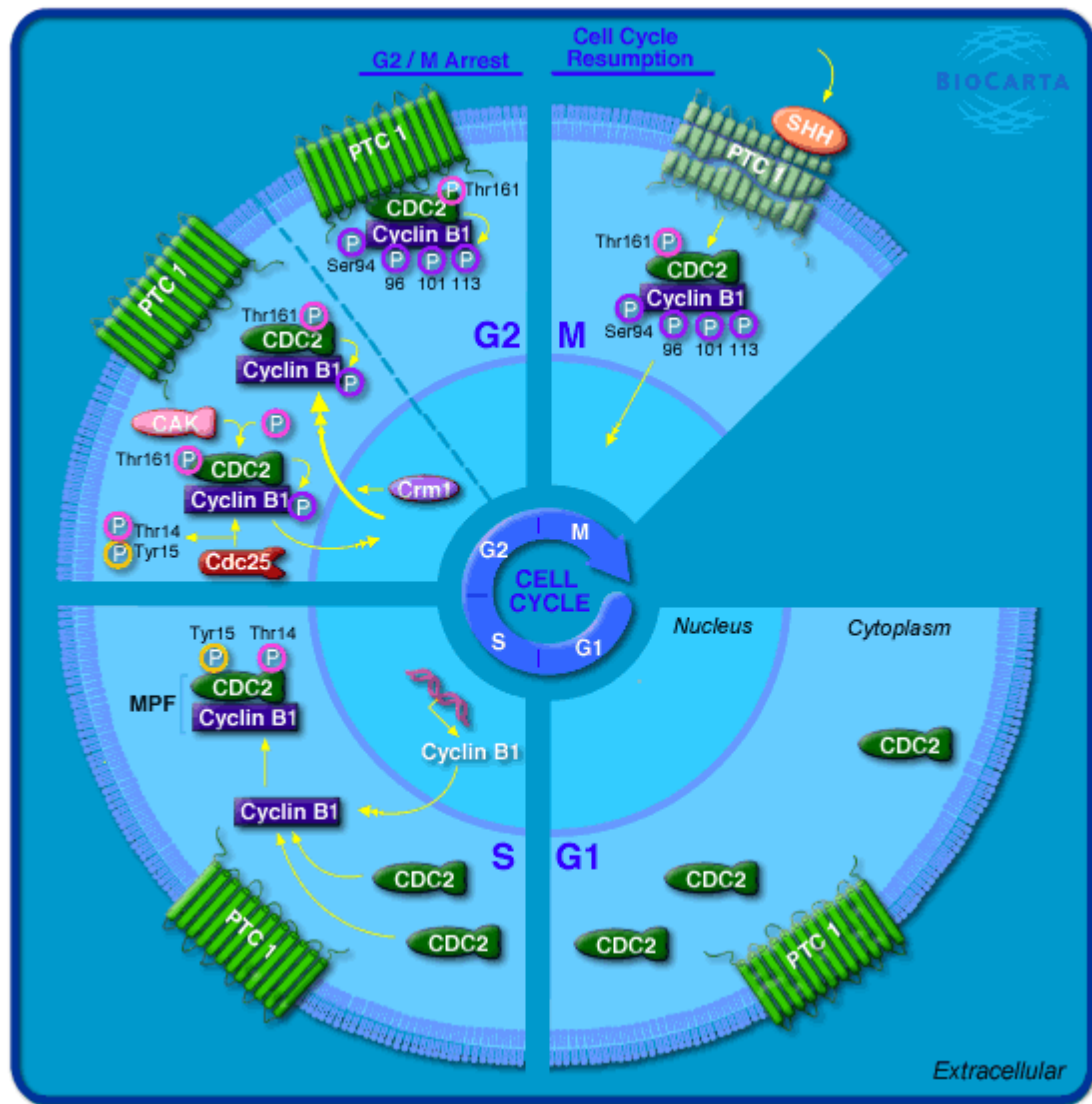


Figure 7.10: BIOCARTA PTC1 Pathway, taken from <http://www.biocarta.com/pathfiles>

The genes included in this pathway are:

CDK7 cyclin-dependent kinase 7

CDK1 cyclin-dependent kinase 1
MNAT1 menage a trois homolog 1, cyclin H assembly factor 1
PTCH1 patched homolog 1
SHH sonic hedgehog homolog
XPO1 exportin 1
CCNB1 cyclin B1
CCNH cyclin H
CDC25A cell division cycle 25A
CDC25B cell division cycle 25B
CDC25C cell division cycle

42 genes are involved in the REACTOME 'Triglyceride biosynthesis' pathway: *ACLY*, *ACSL1*,
ACSL3, *ACSL4*, *ACSL5*, *ACSL6*, *AGPAT1*, *AGPAT2*, *AGPAT3*, *AGPAT4*, *AGPAT5*, *AGPAT6*,
AGPAT9, *ACACA*, *DGAT1*, *DGAT2*, *ELOVL1*, *ELOVL2*, *ELOVL3*, *ELOVL4*, *ELOVL5*, *ELOVL6*,
ELOVL7, *GK*, *GK2*, *GK3P*, *GPAM*, *GPAT2*, *GPD1*, *GPD1L*, *HSD17B12*, *HSD17B3*, *LCLAT1*,
LPCAT1, *LPCAT4*, *LPIN1*, *LPIN2*, *LPIN3*, *FASN*, *SLC25A1*, *TECR* and *TECRL*.

7.4 Discussion

Key findings:

- *ANKRD36C* was significantly associated with unfavourable KD response in the c-alpha gene-based test. *XPO1* reached suggestive significance in uni-directional gene-based tests. Common themes identified amongst genes of interest include protein transport, cell cycling, glucose transport and neurological processes.
- Group 0, 1 and 2 analysis results indicated that non-synonymous variants were mostly driving the association signals and that potential results of interest may be lost when stringent filtering is used.

- The ERBB1 Internalisation pathway reached significance when using the count of case-unique alleles test; other pathways most closely associated with favourable or unfavourable KD response are biologically plausible.

7.4.1 An overview of strategies adopted to increase power

With such a small sample size, association studies are problematic because a variant (or group of variants in a gene) would have to be present in a sufficiently large number of individuals in one response group in order to generate a significant p-value. For this reason, a range of strategies were adopted to increase power.

In addition to selecting individuals with the most extreme (and consistent) KD response, power to detect rare risk variation was increased in this study by combining variants within genes. This approach is particularly important for complex disorders or phenotypic traits, such as the KD, for which the effect of a particular gene, let alone a particular variant, on the phenotype is thought to be modest. The sole effect of such a variant would not be detected, especially when comparing variation across the whole genome [660]. Gene-based tests, however, can still be underpowered to detect an association; power depends on the sample size, the number of variants per gene, variant frequencies, effect sizes and the direction of their effect [963]. Of course, in order to have sufficient power, variants must firstly be present in the same gene [944]. For a specific gene, one will have more power to detect an association if a higher proportion of variants in that gene are causal, if these variants have higher MAFs and larger effect sizes, and (for gene burden tests) if the effects of the variants are in the same direction [1050]. Using the sum of rare variants gene burden analysis, it has been reported that, in a study with 1,000 cases and 2,000 controls, there is 80% power to detect a minor allele with an odds ratio of 3.1 in a gene with 10 variants, each of frequency 0.001 [714]; for genes with 100 variants of the same frequency, there is 80% power to detect minor alleles with an odds ratio of >1.6.

Power also depends on the test used and, for this reason, various gene-based association tests were conducted. The c-alpha test was the first choice, primarily based on the fact that it allows for the effects of variants to be bi-directional. Power was found to be low in the c-alpha test when all variants in a gene were detrimental with regards KD response – in such cases, tests that assume a uni-directional effect of variants generate lower p-values. Furthermore, if results from the c-alpha test were to be used for pathway analysis, it would be difficult to gauge whether genes in that particular pathway affected KD response in a favourable or unfavourable way, or indeed whether they all had the same direction of effect. For these reasons, tests that assume a uni-directional effect of variants were also conducted. The genes of most interest tended to be identified from more than one approach (the single-variant test or one/several of the gene-based tests).

In addition to analyses with no filtering, association tests were conducted with various grades of filtering to exclude variants that were less likely to be of functional consequence. This is based on the premise that, even if a gene harbours phenotypically relevant variation, it may still harbour many other variants that are phenotypically neutral, which would create noise in a gene-based test and mute any association signal. Furthermore, it is difficult to interpret how synonymous variants or those of unknown consequence affect the gene in question. One may conduct a literature search, but this will only help if previous functional studies have been conducted. On the other hand, if these variants are influential then potential true genotypic-phenotypic associations will be lost by excluding them from analyses.

In the gene-based analyses, very similar results were obtained in group 0 and group 1 analyses. This does not necessarily mean that synonymous variants or those of unknown consequence are phenotypically neutral, but it indicates that non-synonymous variants are driving the association signals. Group 2 analyses contained too few variants to generate

low p-values. Even more stringent filters could have been adopted, for example, using previously defined methods to identify high-confidence loss-of-function variants (proposed criteria include destruction of all known protein-coding transcripts of the gene, removal of more than 5% of the remaining protein, and so on) [797, 815] but power would have been even further reduced due to the low number of variants. The results obtained highlight the dangers of missing potentially useful results when adopting stringent filtering criteria, without first considering the potential effects of other variants. For example, one gene of interest, *XPO1*, would not have been identified if only loss-of-function variants had been included in analyses. For this reason, a graded filtering approach was adopted and bioinformatics tools were used post-analysis to predict functional consequence. Adopting stringent criteria to identify truly damaging variants is better suited to Mendelian disorders, where candidate genes tend to be identified by narrowing down variants.

With a complex trait such as KD response, it may not only be different variants within a gene that influence the trait, but also different genes. Variation in a particular gene that influences KD response may be present in only one or a few individuals. Even with a large sample size, the effects of these genes may be difficult to pick up, depending on how many individuals they were present in. For this reason, a self-contained pathway analysis was performed, where variants are pooled into genes and then pooled into pathways. If several genes within a pathway influence the phenotype, this approach should be more powerful than single variant or gene-based approaches. Power depends on the proportion of genes in the pathway that have more variants in one response group. Presumably due to the small sample size, the pathway analysis did not generate significant results, but those pathways with the lowest p-values are certainly of interest.

Even with variant-pooling approaches, the cohort size was too small to generate significant associations from genes or pathways with such (small) effect sizes. One may argue that, in

set-based analyses, adjusting for multiple testing based on the number of sets (whether this be genes or pathways) is overly conservative because many sets do not have sufficient variation to achieve the 'asymptotic properties for the test statistic' [944]; that is to say, if too few variants within a gene/pathway are present in the entire cohort, independent of KD response status, a low or statistically significant p-value could not be generated, unless the cohort size was very large. It has been suggested to adjust for the number of genes/pathways with an i-value below a certain threshold (such as 1×10^{-3}), based on the premise that those with an i-value above the threshold have no power to detect an association. How to determine this threshold, however, is unclear, and this practice has not been readily adopted in publications.

7.4.2 Potential sources of error

As outlined in the Introduction, there are opportunities for error in several steps in the exome sequencing analysis process; false variants in a gene can be called because of errors in alignment or mismatched reads and, if these errors occur at a different rate in cases and controls, this would bias association tests. Some aspects in the sequencing and variant-calling process are out of the researcher's control if these are undertaken elsewhere, whereas other aspects (predominantly post variant-calling) can be manipulated with quality control procedures. In this study, the alignment tool used is considered the most accurate (although also the most time-consuming) of all available and the quality-control criteria used to call variants rendered high-quality performance metrics, indicating that a high proportion of the variants called are true variants. The quality control parameters used by ION Neurogenetics Unit have been developed by UCL genetics workers and optimised over years of study with many different datasets. Of course, there will always be some uncertainty regarding optimal pipeline parameterisations for specific experiments, but the fact that performance metrics were not greatly improved by more stringent quality

control criteria was a good sign; further increasing stringency filters would increase the rate of false negatives.

Two aspects that could potentially be improved would be a) use of multi-sample variant calling instead of single-sample variant calling, and b) use of multiple variant-calling pipelines, instead of just one pipeline.

- a) The advantage of calling variants simultaneously in a group of people is that the genotype at a particular locus for an individual is informed by that of other people: a variant is less likely to be false if it is present in more than one individual. This is particularly beneficial with low-coverage sequencing data. For example, in single-sample calling, one may have little confidence in an SNV with only two supporting reads, but if it is present in many samples, one has the equivalent of more reads. The disadvantage is that more true variants, in particular rare and private variants, will be lost using multi-sample variant calling. Some algorithms may be better suited to multi-sample calling than others: the sensitivity of GATK and glfMultiples were found to be increased with multi-sample calling, whereas SAMtools missed many variants using this approach [979].
- b) Capture varies between bioinformatics algorithms; when using 'near-default' settings, not all variants called from the same raw sequencing data using various algorithms are concordant and 'unique-to-pipeline' variants exist, even in regions of high coverage [1005, 1062]. Concordance between pipelines was found to be lower for indel-calling, compared to SNV-calling. These discrepancies are likely due to the underlying models utilised by the variant-calling algorithms, such as how the prior probabilities are set [1005]. By only using those variants that are called by different pipelines, regardless of parameterisation, the false-positive rate can be minimised, although some unique-to-pipeline true variants may be missed. This

complex approach would require access to various pipelines (not all are freely available) and knowledge of how to use all necessary programmes; it would also be very time-consuming. Samples could be sequenced by various platforms that use different technologies, if the budget allowed for this. Some algorithms can also account for potential batch effects when calling variants [997].

Measures were taken to account for other potential sources of bias in association analyses that were within the researcher's control. For example, several indicators of relatedness due to population specificity or ancestry were used to create a cohort that was as homogeneous as possible: proportion of non-synonymous variation, heterozygosity and pairwise IBS. Visualisation of these parameters allows for exclusion of clear ethnic outliers. One may also use standard deviation from the mean to exclude outliers, but there does not seem to be a consensus regarding what is considered as 'too far' from the mean. For example, outliers have been defined as individuals whose ancestry was at least 2 [1063], 3 [1064] and 6 [1065] standard deviations from the mean.

Other strategies include permutation and inclusion of PCA covariates in association analyses, but even these have not consistently been shown to effectively control for population stratification in rare variant analyses [962]. This is the disadvantage when working with whole exome (or genome) sequencing data: as it is relatively new, approaches to deal with such potential problems as population stratification and software to implement these approaches are not as well-developed as they are for common variant GWAS with array data. Exploration of complex traits with whole exome sequencing data has only recently begun; this approach is more well-established with Mendelian disorders, in which case, instead of association studies, candidate genes tend to be narrowed down by process of elimination according to, for example, novelty and likelihood of functional consequence of variants. With this approach, potential sources of bias such as population

stratification are less of an issue. In any case, it is commonly cited that it is ideal to have a homogenous population before conducting association analyses, whether this be by selective recruitment and/or identification of genetic outliers. Various indicators of ancestral homogeneity have been used in this chapter – all indicate a balance between responders and non-responders and so bias due to population stratification should not be an issue. This was supported by the fact that λ_{GC} was equal to 1.00 in the single variant test.

Another potential source of bias in association studies with sequencing data is variability in coverage and variant discovery between cases and controls. This can result in bias similar to that from population stratification if such errors are not balanced between cases and controls [660]. Identification and exclusion of outliers with a lower total number of called exonic variants than what is commonly expected from whole exome sequencing, or with excess or reduced proportions of singleton variants, exonic bases with X-fold coverage, mean Ti/Tv ratio and genotype quality scores, compared to all other individuals, ensured a homogenous cohort. Such approaches further limit the sample size and thus the power to detect an association, but they are a necessary step in reducing bias and ensuring confidence in the results obtained. Even though coverage in this cohort was not as high as in other published cohorts (for example, with the 1000Genome Project, all exome samples had 70% of their targets covered at least 20x

[<http://www.1000genomes.org/category/sample>]), it was uniform and so should not have biased association studies. In the future, funds should be directed towards sequencing more samples, even if this is at the expense of coverage, rather than sequencing fewer samples at higher coverage, in order to increase power. Alignment of reads has the greatest effect on proportions of true/false variants and, even if samples are sequenced to a greater depth, this does not necessarily lead to detection of more variants (personal communication with Dr Alan Pittman).

The homogeneity of the cohort in terms of performance metrics is also an indication of a lack of batch effect. This would not necessarily identify variability in specific subsets of genes, which may be differently affected by batch effects [1066] but it shows that, in general, the data is of high quality, independent of batch or other factors, such as case/control status. Furthermore, all individuals were sequenced on the same machine and samples were handed over to the ION Neurogenetics Unit with responders and non-responders randomly ordered.

The possibility cannot be discounted that non-coding variation causes variability in KD response, either separately or in conjunction with coding variation. Whole genome sequencing would allow for investigation of this, although interpretation of non-coding variants, particularly intergenic variants that cannot be easily assigned to a gene, is problematic. Relevant large indels or structural variants may have been missed by sequencing [516] and, if one wanted to concentrate on such variation, perhaps a combination of sequencing- and array-based methods would be appropriate [994].

7.4.3 Results of interest

7.4.3.1 Single variant and gene-based approach

The only statistically significant result was for *ANKRD36C* from the c-alpha test. This result seems to be driven by rs5005869 and other variants that have a large difference between minor allele counts in responders and non-responders; there are many other variants within *ANKRD36C* with less of a difference between responder/non-responder alternative allele counts and so this gene was not among those with the lowest p-value in the tests that assume a uni-directional effect of variants. The difficulty lies in a) interpreting the functional consequences of the variants of most interest on the gene, and b) interpreting how this may influence KD response. There is little information available on the function of *ANKRD36C*. It has the GO term 'ion channel inhibitor activity' – identified from

haemangioma cells from an individual with oncogenic osteomalacia and hypophosphataemia, it was found to inhibit renal epithelial phosphate transport [1067]. Mutations in other ankyrin repeat proteins have been associated with cancer and bare lymphocyte syndrome, which causes immunodeficiency [1068]. According to public databases, *ANKRD36C* is most highly expressed in the brain (<http://embl-ebi.org>) and it interacts with PRKACB and UBC proteins (<http://thebiogrid.org>). *PRKACB* encodes a subunit of cAMP-dependent protein kinase, through which cAMP exerts its effects, for example, inhibiting gluconeogenesis, stimulating lipid oxidation and regulating apoptosis [1069, 1070]; *UBC* encodes ubiquitin C, which is involved in many processes, including cell cycle progression and apoptosis [1071].

Other genes with the lowest p-values in gene-based tests, or with the largest difference between responder/non-responder alternative allele counts, are easier to interpret. For example, mutations in *L2HGDH* have been associated with L-2-hydroxyglutaric aciduria, characterised by mild to moderate psychomotor retardation, cerebellar ataxia, variable macrocephaly, and epilepsy [1072]. This is a recessive disorder and so, considering that the genotype of the individuals with the putatively damaging variant in *L2HGDH* was A/C, this would not lead to a diagnosis of a neurometabolic disorder; rather, this indicates that the function of the L-2-hydroxyglutarate dehydrogenase enzyme may be partially affected – not necessarily enough to result in a neurometabolic phenotype but enough to influence KD response. Alpha-ketoglutarate is the keto acid produced by the breakdown of the excitatory neurotransmitter glutamate (via the intermediate 2-hydroxyglutarate), which then enters the citric acid cycle to yield ATP. Individuals with a damaging variant in *L2HGDH* may not be able to optimally yield energy from glutamate. Alpha-ketoglutarate is transaminated back to form glutamate, which can be used to form the inhibitory neurotransmitter GABA. Reduced GABA levels or a suboptimal balance of glutamate and

GABA levels would influence neuronal excitability and potentially prevent the KD-induced increased seizure threshold. As outlined in Chapter 1, it has been suggested that the KD reduces seizure frequency by increasing the availability of glutamate, leading to elevated levels of GABA to dampen hyperexcitability in the brain, and also that the KD replenishes citric acid cycle intermediates that are made deficient by seizures. Based on these theories, mutations that hinder the glutamate/alpha-ketoglutarate/glutamine/GABA cycle would thus likely lead to an unfavourable response to KD.

XPO1 (which included a variant unique to 13 non-responders that was predicted to be damaging by all algorithms available) encodes exportin 1, which mediates nuclear export signal-dependent protein transport. This protein, first associated with nuclear export of human immunodeficiency virus Rev protein [1073], has been shown to be important for the stress response. Exportin 1 has since been implicated in recovery from stroke, as it enables HMGB1 (released from astrocytes in response to inflammatory cytokines) to translocate from nucleus to cytoplasm before release [1074]. MAPK ERK was also involved in the signalling process and its inhibition decreased HMGB1 upregulation in the stimulated astrocytes. The stress-induced translocation of MK5 (MAPK-activated protein kinase 5) from the nucleus to the cytoplasm has been shown to be dependent on exportin 1 [1075], as has localisation of PAPA1 (which activates expression of many genes induced by oxidative stress) in response to oxidative stress [1076]. Exportin 1 has also been shown to play a role in apoptosis: inhibition of XPO1-mediated nuclear export induced rapid apoptosis in leukemic CD34(+) progenitors [1077]; ablation of the nuclear export signal to which exportin 1 binds in the second SH3 domain of Crk II increases the proapoptotic activity of the Crk protein [1078].

Exportin 1 may influence glucose homeostasis via translocation of hepatic glucokinase (in the postprandial state, the simultaneous rise in glucose and insulin increases glucokinase

activity; in the fasting state, the combined decrease in glucose and insulin concentrations and increase in glucagon concentrations, stop glucokinase activity [1053]). It has been reported that, when intracellular glucose concentrations are high, glucokinase translocates from the nucleus to the cytoplasm via exposure of its active nuclear export sequence [1054]. Others have found that glucokinase regulatory protein mediates the nuclear export of glucokinase, as well as its import, which does not involve interaction with exportin 1 [1079].

The theme of glucose metabolism links with the functions of other genes of interest identified from single variant and gene-based tests. Genes may share similar functions and may (jointly or independently) influence KD response, even if they exert their effects in different areas of the body. For example, while *XPO1* may influence glucose metabolism via an effect on glucokinase in the liver, *GBF1* may exert an effect through activation of ARF5, which is thought to regulate insulin-stimulated translocation of GLUT1 to the cell membrane in adipocytes [1080]. *TRPM5* is expressed in pancreatic β cells and is also thought to play a role in glucose-stimulated insulin secretion [1052, 1081]. As glucose enters pancreatic β cells and is metabolised, the ATP-to-ADP ratio is increased and K_{ATP} channels close. This depolarises the membrane and causes an influx of calcium through voltage-dependent calcium channels. In theory, this then leads to insulin release, although K_{ATP} channel closing alone has been found to be insufficient to depolarise the membrane to the extent needed for insulin release to occur [1082]. Activation of *TRPM5* and subsequent depolarisation of the plasma membrane may help depolarise the membrane to an extent sufficient for insulin release. It is thus feasible that variation in *TRPM5* that affects the role of the protein in glucose-stimulated insulin release may influence KD response; *TRPM5* is also expressed in the brain and may influence neuronal glucose metabolism (for example, the preference of neurons to metabolise glucose or ketone bodies) or it may play a more

indirect role by affecting glucose uptake elsewhere in the body. The extent to which variation influences KD response may depend on other genetic variation, for example, in *KCNJ11*. An alternative theory is that variation in *TRPM5* affects people's food preferences (*Trpm5* knockout mice of the strain are unable to detect sweet taste [1083]) and so certain individuals may be more or less likely to be non-adherent to the diet. Others have found that reduced glucose intake in *Trpm5* knockout mice is independent of taste [1084]. As the TRPM5 channel is also present in the gastrointestinal tract, the authors suggested that a lack of preference for carbohydrate may be due to selective blockage of glucose utilisation and the subsequent dopamine reward signals.

Other themes identified among the genes of most interest may also be relevant to KD response. GTP/GTPase-activating proteins and GTP-binding proteins have been shown to influence insulin signalling and glucose homeostasis [1085-1087] and apoptosis [1088-1090]. The genes associated with Alzheimer's disease (*APBA2* and *GBF1*) are also involved in several processes, such as neurotransmitter release (*APBA2*) [1091], stabilisation of amyloid precursor protein (*APBA2*) [1092] and maintenance of Golgi structure and function (*GBF1*) [1093]. The latter function of *GBF1* is impaired in situations of metabolic stress that deplete ATP, such as glucose deprivation [1094]. It is biologically plausible that such genes, which have multiple functions and/or are affected by various factors, influence KD response, but it is unknown exactly how. Genes may influence KD response via different mechanisms in different people, depending on other genetic and environmental factors.

Protein binding/transport was another common theme amongst genes of interest, but the general nature of this term means that it does little to enhance understanding of how such processes influence KD response.

7.4.3.2 Pathway-based approach

Despite the fact that several genes with the lowest p-values or the greatest difference in case-control alternative allele counts are of interest, an examination of allele counts reveals that extreme KD response in this cohort is not influenced by one single gene – not surprisingly, due to the complex nature of the trait. Thus, rather than looking at a list of genes of interest, it makes sense to concentrate more on pathways that may influence KD response. Some genes within these pathways may have more of an effect on the phenotypic trait than others. The pathway-based approaches have highlighted some common pathways and themes that may potentially influence KD response, namely stress response, neurological processes and lipid metabolism.

The ERBB1 (human epidermal growth factor receptor) pathway was significantly associated with unfavourable KD response in the case-unique alleles test. ERBB1 lies on the cell surface and, when activated by ligand-binding, endocytosis occurs: the receptor internalises and moves to the early endosomal compartment of the cell. Depending on which ligand has bound to ERBB1, the receptor may be degraded or recycled and moved back to the cell surface [1124], and this has a downstream effect on several cascades (either diminishing or prolonging their signals [1125]), including the MAPK, AKT and stress-activated protein kinase pathways, which lead to cell cycle progression, growth, survival and proliferation [1126]. Mutations that lead to *ERBB1* overexpression have been implicated in the pathogenesis of many cancers [1127], potentially because this leads to constant activation of the receptor and thus uncontrolled cell division. *ERBB1* is expressed in the brain, including in neurons and astrocytes, and mice with mutations in *ERBB1* that survive display progressive neurodegeneration characterised by apoptosis and upregulation of *c-fos*, which is also thought to act as a mediator of cell death [1128]. ERBB1 overexpression has frequently been found in many tumours, including in the brain [1129]

and ERBB1 inhibitors (gefitinib and erlotinib) are used to treat non-small cell lung cancer. ERBB1 ligands, such as heparin-binding epidermal growth factor-like growth factor and transforming growth factor- α , have been shown to promote survival of neurons and axon regeneration in response to induced seizures or brain injury and have thus been proposed as neuroprotective agents [1130-1132]. Depending on which components of the pathway are affected (other ERBB1 ligands or other members of the ERB family may compensate for impaired function of ERBB1 or vice versa), KD response may be adversely affected by an impaired stress/apoptotic response to seizures. One may draw a parallel with response to gefitinib, which is thought to be mediated by AKT signalling, which has antiapoptotic properties: mutations in the ERBB1 pathway (not just ERBB1 itself) lead to reduced internalisation rates and unrestrained signalling of AKT, thus reducing apoptosis and allowing proliferation [1133]. It is also of interest that *ERBB1* expression was found to be lower in adipose tissue of insulin-resistant or diabetic obese women, compared to those with normal insulin sensitivity, and that inhibition of ERBB1 prevented adipogenesis in human preadipocytes [1134]; the authors raised the possibility that reduced *ERBB1* expression is, akin to insulin resistance, a way of preventing further dyslipidaemia and insulin resistance by inhibiting lipolysis. An ERBB1 inhibitor induced an improvement in glucose tolerance and insulin sensitivity in mice, and a reduction in circulating levels of the cytokines IL6 and TNF- α , consistent with reduced inflammation [1135].

The PTC1-signalling pathway reached suggestive significance in the case-unique alleles test, when testing for association with unfavourable KD response. *PTC1* encodes patched 1, a sonic hedgehog (SHH) receptor that represses SHH cell signalling, affecting different phases of the cell cycle. When bound to SHH, patched 1 releases the transmembrane protein smoothened, which leads to activation of the GLI transcription factors, which control the transcription of hedgehog target genes, allowing progression to the G1 phase of the cell

cycle. The G2/M (Gap 2 – the last stage of interphase – to mitosis) transition phase of the cell cycle is also affected: when unbound, patched 1 binds to cyclin B1 and CDC2, which blocks M-phase promoting factor activity and progression through the G2/M transition phase; when bound to SHH, PTC1 is degraded and so cannot bind to cyclin B1; cyclin B1 then enters into the nucleus

(http://www.broadinstitute.org/gsea/msigdb/cards/BIOCARTA_PTC1_PATHWAY). Exportin 1 is also involved in this pathway, as it mediates the localisation of the cyclin B1/CDC2 complex and thus the entry of cells into mitosis [1095]; binding to exportin 1 causes nuclear export of the cyclin B1/CDC2 complex and disruption of this binding leads to an accumulation of this complex in the nucleus.

Both PTC1 and SHH have been shown to play an important role in both prenatal and postnatal brain development through the proliferation of neural progenitors and determination of cell fate [1096]. For example, Shh was shown to maintain the survival of neuroepithelial cells by preventing Ptc1-induced apoptotic cell death [1097]. Mutations in *SHH* are the most common cause of holoprosencephaly, a disorder in which the embryonic forebrain does not divide into two cerebral hemispheres [1098, 1099]. Holoprosencephaly is associated with a spectrum of symptoms, ranging from serious malformations of the brain, facial abnormalities and epilepsy, to microcephaly, learning disabilities and motor impairments. Mutations in *PTCH1* have also been associated with a holoprosencephaly-like phenotype [1100-1102], as well as Gorlin syndrome, characterised by developmental malformations and a high predisposition to skin basal cell carcinoma development [1103, 1104]. Methylation of *PTCH1* and loss-of-function mutations in the gene have been associated with tumour development [1105-1107]. Mutations in genes encoding several components of the SHH pathway, with the overall effect of increasing SHH activity, have

been associated with skin tumours [1108-1110]. It is predicted that mutations in this pathway that lead decreased SHH activity may cause holoprosencephaly [1111].

The SHH signalling pathway has been implicated in the process of hippocampal neurogenesis after injury and increased mRNA levels of Shh and Gli1 were found in hippocampus immediately after ischaemia [1112]. Hypoxia/ischaemia has also been demonstrated to induce expression of Shh in cardiomyoblasts, astrocytes, neurons, neural progenitor cells and testicular tissue [1113-1117]. The hypoxia-induced increased expression of SHH is dependent on hypoxia-inducible factor-1 α (HIF1 α) [1114, 1116]. It is unknown exactly how the PTC1/SHH signalling pathway exerts a neuroprotective effect, and there are several possibilities. SHH has been shown to activate the PI3K/AKT pathway, upregulate *BAD* expression and inhibit the ERK pathway, thus decreasing cell apoptosis in neurons [1118]. The inhibition of Smo and subsequent activation of Gli1 are thought to be involved in the neuroprotective mechanisms of SHH but it is unknown how they exert these effects [1119]. SHH has been shown to have neurotrophic effects in midbrain dopaminergic neurons but not in motor neurons [1120]. The authors found that this neurotrophic effect could not be attributed to the stimulation of proliferation by SHH, as the neurotrophic activity was on post-mitotic neurons rather than on dividing progenitor cells. Others have shown SHH gene expression to be induced by BDNF in rat neurons [1121] and it has been suggested to investigate whether SHH-mediated neuroprotection involves BDNF [1119]. Children with autism were found to have higher serum levels of SHH and BDNF compared to age- and sex-matched controls [1122].

It is biologically feasible that mutations in the PTC1/SHH pathway adversely affect KD response as the neuroprotective response to seizures is impaired. This may be due to impaired regulation of cell apoptosis, neurotrophin release and/or response to oxidative stress. Considering that the starvation stress response has been correlated with the

oxidative stress response in yeast [1123], the KD may result in similar changes in gene expression as those caused by hypoxia (increased expression of components of the PTC1/SHH pathway, with a neuroprotective effect). Individuals without damaging variants in the various components of this pathway may respond better to the KD as this neuroprotective effect is not impaired. These mechanisms may not only mediate the antiseizure effects of the KD, but also the neuroprotective or anticarcinogenic effects of the KD for other neurological diseases and cancer.

The pathway most closely associated with favourable KD response was triglyceride biosynthesis. This pathway consists of fatty acyl-CoA biosynthesis, conversion of fatty acyl-CoA to phosphatidic acid, conversion of phosphatidic acid to diacylglycerol, and conversion of diacylglycerol to triacylglycerol. Fatty acids may be synthesised from dietary fat intake or de novo synthesis from acetyl CoA (mostly obtained from conversion of pyruvate [from glucose] to acetyl CoA in the mitochondria) [1136]. As part of the wider pathway, 'Fatty acid, triacylglycerol, and ketone body metabolism', genetic variation in the 'triglyceride biosynthesis' pathway may indirectly influence ketone body production and metabolism. Hydrolysis of triglycerides generates fatty acids for use as an energy substrate in muscles and other tissues, and for conversion into ketone bodies, as well as glycerol for gluconeogenesis. Variation affecting this pathway may impair or increase the amount of ketone bodies available for brain metabolism, which may influence seizure control (whether this is mediated by a direct or indirect effect, or both, is unknown).

The effects of insulin on triglyceride turnover are also worth noting. Insulin stimulates cells to preferentially use glucose instead of fatty acids as an energy source and it controls triglyceride synthesis in adipose tissue and muscle by increasing glucose uptake and regulating its conversion to glycerol-3-phosphate via glycolysis [1137]. Both acutely and chronically elevated levels of free fatty acids cause reduced glucose uptake and increased

insulin secretion, ultimately leading to insulin resistance and potentially type II diabetes [1138]. Increased serum fatty acid levels are a hallmark feature of KD-feeding, which is needed for ketone body synthesis, to provide energy for the muscles and liver, and to avoid lean body mass breakdown. Severe insulin resistance would not be desirable when following a KD, as some glucose is still needed by the brain. An inability to esterify fatty acids back to triglycerides in the triglyceride-to-fatty-acid cycle would further exacerbate insulin resistance and may also cause excessive concentrations of ketone bodies and thus ketoacidosis.

It is biologically feasible that variation in these pathways influence KD response, although the extent to which KD response is affected will be determined by how damaging the variants are, whether an individual is homozygous or heterozygous for the alternative allele, which genes the variants are present in and, potentially, other environmental factors. For example, the wide phenotypic variability in individuals with holoprosencephaly caused by heterozygous mutations in SHH suggests that the extent to which cortical development is impaired depends on other factors (these may be genetic or environmental) that interfere with the already impaired SHH signalling. Dietary intake of fat and carbohydrate may influence to what extent variation in the triglyceride biosynthesis pathway affects KD response. Depending on how many genes within a pathway harbour damaging variants, individuals may be able to synthesise enough triglycerides and obtain sufficient amounts of fatty acids and ketones for seizure control.

7.4.4 Conclusions and future work

A combination of single variant, gene-based and pathway-based tests have revealed several genes and biological pathways that could feasibly account for variability in KD response in this cohort. Predominant themes include apoptosis/stress response, glucose homeostasis and lipid metabolism.

The difficulty resides in translating these results into clinical practice. Given the complex nature of the trait, in the absence of *SLC2A1* mutations, KD response may not be influenced by a single gene; it may not even be influenced by a single pathway. There may be a different genetic reason for (un)favourable response in different individuals and so no signal would be picked up in an underpowered association study. If this is the case, a larger sample size may indicate several pathways of interest, but only stringent stratification of the cohort (for example, by extent of KD response or presence of certain genetic variants) could help predict which mechanisms are relevant for which individuals. Furthermore, as demonstrated by *XPO1*, a single gene, or a single pathway, may not just influence one process. This overlapping complicates the interpretation of results and further investigation would be required in order to establish exactly how the variants/genes of interest influence KD response and whether the mechanisms differ between individuals.

The problem with investigating the genetic basis to complex traits is that, if phenotypic variability is caused by a collection of variants or genes, one may be limited by the cost of follow-up studies, such as targeted sequencing of genes of interest, or functional studies of specific variants (such as c.T593G:p.V198G in *L2HGDH* or rs79074863 in *XPO1*) that would help elucidate the consequences of such variants on the gene and the exact mechanisms by which this affects KD response [945, 1033]. The first step would be to identify those variants of most interest (variants that are predicted to be damaging in genes with a feasibly relevant role, or several variants within feasibly relevant genes) and confirm them with Sanger sequencing, as is standard practice with variants identified from next-generation sequencing. The variant/gene(s) in question may not be relevant for every individual in the cohort, but it may help to explain variability in KD response in a subset of patients.

8 Conclusions, study limitations and future work

8.1 Conclusions

The aim of this study was to determine whether genetic variation affects response to the KD, with the hope of identifying the biological underpinnings of the antiepileptic effects of the KD and potentially improving targeting of treatment. This is the first exploration of the genetic contribution to KD response and a range of approaches were adopted, including candidate gene sequencing, whole-genome SNP array genotyping and whole exome sequencing. Analyses were underpowered, but similar themes were identified from common and rare variant association analyses. Gene- and pathway-based analyses provided the opportunity to detect complementary information that would not have been easily observed from single SNP analyses alone.

8.1.1 Comments on phenotypic results

KD response can be defined in various different ways. Firstly, when used to treat drug-resistant epilepsy, seizure reduction is commonly used as the primary measure of response. However, as demonstrated with this cohort, other benefits may be obtained from the KD that cause some individuals to continue treatment, despite no (or very little) reduction in seizure frequency. These include seizure-related benefits, such as reduced seizure severity and fewer hospital admissions, and non-seizure-related benefits, such as improved alertness, cognition and sleep. In this study, only seizure frequency was used to classify KD response, as this was the only factor that was systematically commented on in clinic letters and thus could be semi-quantified. A $\geq 50\%$ reduction in seizure frequency is considered clinically meaningful in such drug-resistant cohorts, although other 'stricter' definitions may be adopted to define extreme response, such as $\geq 75\%$ or $\geq 90\%$ seizure reduction. Seizure diaries would likely have to be used in order to treat seizure frequency as a continuous

variable, or to very precisely define seizure reduction (for example, to differentiate between those individuals who achieved $\geq 90\%$ seizure reduction and those who were seizure-free).

Secondly, KD response may be classified at specific points in time or using longitudinal data. This is independent of whether multiple factors, or solely seizure frequency, are used to define response. As shown in this cohort, response is not always consistent over time (seizures may be temporarily exacerbated, for example, due to illness, or for unknown reasons) and so, by defining responders as those with a sustained reduction in seizure frequency, one can be more certain that seizure reduction was due to KD-feeding and not due to chance alone. How long seizure reduction should be maintained for is debatable. On the other hand, it may be useful to conduct analyses taking response at specific time points, as certain genes may have different effects in the short- and long-term. The difficulty with this approach is that it is unknown whether the KD induces different changes in gene expression at the follow-up points used in this study (3-, 6-, 12-, 18- and 24-months+ from commencement of treatment). Obviously, changes associated with starvation have not been assessed for this time period; studies regarding prolonged calorie restriction (that is to say, over months/years rather than weeks) tend to concentrate on clinical measures, such as weight loss/maintenance, insulin sensitivity or lifespan, and studies that have examined gene expression changes focus on one particular time point, or do not specify for how long subjects/animals have followed a calorie-restricted diet. The benefit of the strategy employed with the rare variant array data and exome sequencing data is that results would be the same if analyses were conducted with 3-month response, as responders had sustained seizure reduction at every time point.

8.1.2 Comments on genotypic results

Despite the fact that the recruited cohort was relatively small, several variants/genes were highlighted in this work as being potentially relevant to KD response. rs12204701 reached significance in the 3-month response GWAS, but did not even reach suggestive significance in the summary diet response GWAS. One can only presume that this variant influences its closest gene, *CDYL*, but, even if this were the case, it is unknown how this may influence KD response. rs12204701 may affect expression of other genes. *CDYL* was not one of the genes highlighted from gene-based analyses, either from the chip or exome sequencing, which assigns less credence to the importance of this gene.

The significant association of *ANKRD36C* with extreme KD response in the c-alpha exome sequencing test seems to have been predominantly driven by specific variants, such as rs5005869 (present in 17 non-responders and 1 responder), rs1819101 (present in 12 non-responders and 2 responders) and rs5005868 (present in 6 non-responders and 0 responders). Many other variants within this gene had a lesser difference in non-responder/responder alternative allele counts and, for several variants, more responders harboured the alternative allele than non-responders, which is why this gene did not generate a low p-value in the burden-based tests. The variants with the greatest difference in non-responder/responder alternative allele counts were not consistently predicted by various algorithms to be damaging and so it is difficult to know whether they are relevant or not. No variant typed on the SNP array was found to be present in *ANKRD36C*. Furthermore, only scarce information regarding *ANKRD36C* is present in the literature and so, even if these variants were to influence KD response, it is unknown exactly how this would occur. It is of interest that *ANKRD36C* is expressed in the brain and that it interacts with proteins that influence glucose and lipid metabolism and apoptosis.

XPO1 was the only gene that reached suggestive significance or generated the lowest p-value in gene-based analyses from the chip (3-month response GWAS) and exome sequencing data. This suggests that it may be relevant for KD response, whether it be an extreme response or just $\geq 50\%$ seizure reduction. rs79074863, present in *XPO1*, along with c.T593G:p.V198G in *L2HGDH* are of most interest from the exome sequencing data as they were predicted to affect gene function by all available algorithms, which assigns more credence to the relevance of these genes to KD response. *L2HGDH* did not generate as low a p-value as *XPO1* in gene-based tests because *L2HGDH* also contained other variants that did not have such a differential distribution between responders and non-responders. Neither of these variants was typed on the chip, but the gene-based test indicated that other typed variants within *XPO1* may also be relevant, or that these variants may tag a susceptibility variant (potentially rs79074863). As mentioned in the previous chapter, the biological functions of *XPO1* and *L2HGDH* make them plausible candidates for influencing KD response, although, at least with *XPO1*, they may exert effects through several pathways.

Even with a relatively small sample size, a gene that strongly influences KD response (for example, with variants in the majority of non-responders but few or no responders) should have been identified from association analyses, or examination of responder/non-responder allele counts. However, with the exception of specific metabolic disorders where the genetic basis for KD response is known, the KD is a complex trait and is unlikely to be influenced by a single gene. Genes that influence KD response are likely to have a relatively small effect size – smaller than what the present analyses were powered to detect. This is precisely what this study has found: association analyses highlighted many variants or various variants within genes with a different distribution (some to a greater extent than others) between KD responders and non-responders, many of which could feasibly

influence KD response. It may be that, with a larger sample size, particular genes will be shown to be more relevant than others. For example, if more individuals were exome sequenced and more non-responders (and no or very few responders) were found to harbour damaging variants in *XPO1* or *L2HGDH*, sequencing of these genes could be used as part of the screening criteria for contraindications to KD response. A larger cohort would assign more credence to the importance of these genes and increase confidence that certain results did not occur by chance alone. This could then be extended to other genes that, if their function were compromised (but individuals may still be asymptomatic), may be a contraindication for starting KD treatment, such as those involved in fatty acid oxidation.

Due to the complex nature of KD response and the fact that many genes may account for phenotypic variability, it seems prudent to examine common themes highlighted from the genes of most interest, as well as from pathway analyses. Several common areas of interest were identified, both from common and rare variation analyses:

- i) cell cycle progression and apoptosis
- ii) glucose homeostasis
- iii) neurological processes
- iv) triglyceride biosynthesis

At first sight, these themes appear diverse, supporting the premise that the KD relies on various mechanisms of action. However, these processes have a point of convergence: insulin/insulin-like growth factor 1 (IGF1) signalling and downstream effects on PI3/AKT/mTOR and MAPK/ERK pathways; they either begin with these signalling cascades or interact with one or more of their components downstream.

The PI3/AKT/mTOR and MAPK/ERK pathways are two major pathways controlling the metabolic and mitogenic roles of insulin/IGF1. Signalling through these pathways is initiated by stimuli from growth factors that bind receptors in the cell membrane, such as insulin-like growth factor 1 receptor (IGF1R), platelet-derived growth factor receptor (PDGFR), epidermal growth factor receptor (EGFR) and human epidermal growth factor receptor 2 (HER2) (<http://www.sabiosciences.com/pathway>, [1139]). As outlined in Figure 8.1, activation of insulin-receptor substrate 1 (IRS1) (this may be from phosphorylation of IGF1R by IGF1 or insulin, or phosphorylation of the insulin receptor by insulin or IGF1, which allows the receptors to bind to IRS1) leads to a cascade of events that has many implications: via phosphorylation of RAS and then RAF and MEK, MAPK and ERK are activated, and subsequently ERK1 and ERK2; through activation of PI3K, AKT is activated; both ERK and AKT can activate mTOR; AKT can also inhibit AMPK, which further stimulates mTOR activity.

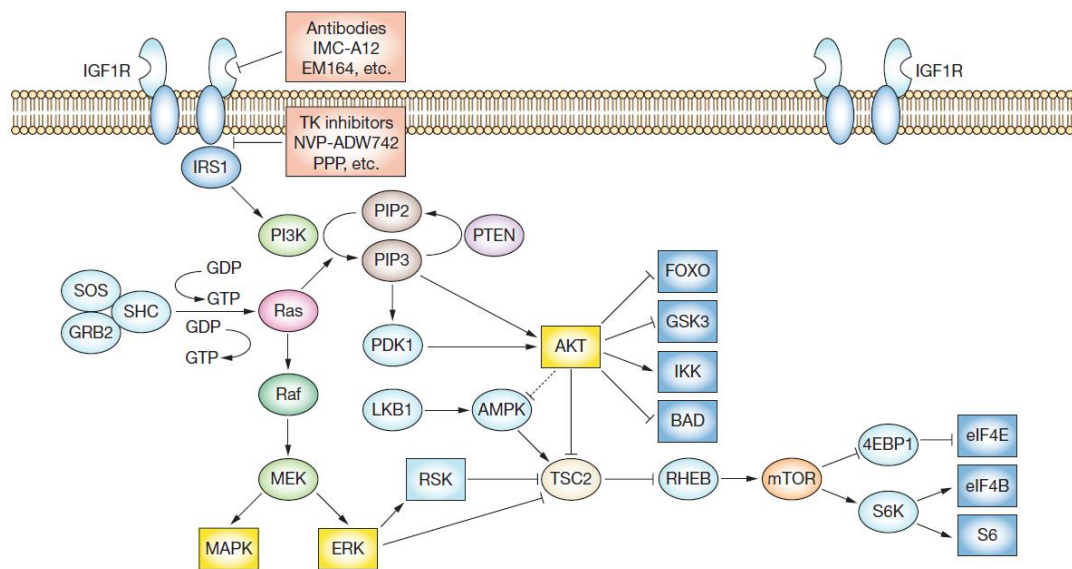


Figure 8.1: Diagram of the insulin/insulin-like growth factor 1/insulin receptor substrate signalling cascade, taken from Tao et al. 2007 [1140]

This has the following effects:

Activation of MAPK/ERK: mediates cell growth, survival and differentiation, predominantly through regulating transcription of genes such as c-fos. Whether activation of this pathway promotes or protects against neuronal apoptosis most likely depends on the stimulus, cell type and the duration of ERK activation [1141].

Activation of PI3/AKT: AKT regulates cell growth, survival and cell cycling through phosphorylation of a number of substrates. For example, by inhibiting the pro-apoptotic molecules BAD and FOXO, it mediates cell survival. AKT has been shown to regulate proteins involved in neuronal function, such as GABA_A receptor, ataxin 1, and huntingtin (http://www.cellsignal.com/reference/pathway/Akt_PKB.html) and thus plays a role in synaptic plasticity and neurodegeneration. AKT is also involved in insulin signalling and glucose metabolism: for example, in the liver, FOXO1 activates transcription of genes encoding gluconeogenic enzymes and AKT suppresses this; AKT is also involved in the translocation of GLUT4 to the cell membrane in response to insulin [1142] and suppresses β -oxidation of fatty acids [1143].

Through activation of mTOR, either by ERK or AKT stimulation or lack of inhibition from AMPK, cell growth and proliferation are positively regulated through the promotion of protein and lipid synthesis, the upregulation of key glycolytic enzymes and the inhibition of catabolic processes such as autophagy [1144, 1145].

There is substantial cross-talk between kinases of the PI3K/AKT/mTOR and Raf/MEK/ERK pathways, and inhibition of one cascade activates the other [1139].

Insulin/IGF1 and its downstream pathways have received much attention with regards to their role in cancer and the KD. Expression of IGF1R has been shown to be higher in cancer cells, compared to normal tissue [1146, 1147], and genetic variation in IGF1 has been associated with increased cancer risk and poor survival rates [1148-1150]. The insulin/IGF1

pathway influences many other pathways, such as epidermal growth factor receptor and HER2, which have both been targeted to enhance apoptosis and reduce cellular invasion and angiogenesis in tumours [1140, 1151]; resistance to drugs that target these molecules may be prevented by treatments that inhibit IGF1R signalling [1152-1155]. In light of this, it is feasible that these pathways underline the mechanisms of action of the KD for cancer. Plasma IGF1 levels reduce in correlation with plasma glucose levels and both are predictive of brain tumour growth [1156]. This is not surprising, considering that tumour cells cannot metabolise ketone bodies and so the KD (or indeed, starvation or calorie restriction), in effect, starves the tumour, and inhibits its growth. However, the effect of the KD may be wider than simply metabolically isolating tumours: it may enhance apoptosis (calorie restriction and the glycolysis inhibitor 2DG have been shown to enhance apoptosis [1157, 1158]), reduce inflammation (calorie restriction reduced the expression and activity of pro-inflammatory markers in brain tumours [1159]) and reduce angiogenesis (vascularity) [1158].

These pathways have also been implicated in the beneficial effects of calorie restriction (without malnutrition) on ageing. Reduced insulin/IGF1 signalling through deletion of the insulin receptor, *IGF1R* or growth hormone has been associated with longevity in animal and human studies [1160]. It is thought that the metabolic shift associated with low glucose levels and thus a low insulin drive may influence longevity in various ways, many of which correspond to those implicated with cancer. These include suppression of NF κ B and subsequent reduced activation of *COX2*, *TNF α* and other pro-inflammatory genes and antiapoptotic molecules, upregulation of several antioxidant enzymes: and activation of *PPGC1 α* (which leads to increased oxidative stress resistance), and increased mitochondrial biogenesis (which leads to less ROS production as mitochondrial workload is less) [1160, 1161]. The degree of calorie restriction and diet quality are also likely to play a role [1162].

Might the KD alter insulin/IGF1 signalling and this is how it exerts its antiseizure effects?

The KD has been shown to increase insulin sensitivity in animal studies and in humans [177, 192, 416, 1163-1165] and so these pathways must be altered in some way. Genes and pathways related to cell cycle progression, apoptosis and glucose metabolism, which interact with or can be linked back to the insulin signalling cascade, were associated with unfavourable KD response in exome sequencing analyses (the direction of effect was unknown in GWAS analyses, although the same themes were highlighted); it is unknown whether variation inhibits or enhances the activity of certain genes. It is feasible that increased insulin/IGF1 signalling may reduce seizures, potentially due to the effects of downstream pathways involved in either cell survival/apoptosis or metabolism. Insulin and IGF1 receptors are present in most brain regions and increased expression of IGF1 and IGF binding proteins occurs in response to ischaemia in rats [1166]. Insulin has been found to reduce neuronal excitability, potentially by activating GABA channels [1167], and to inhibit hippocampal epileptiform-like activity by activating K_{ATP} and large-conductance, calcium-activated- K^+ channels via an MAPK-driven process [1168, 1169]; insulin and IGF1 regulate neurotransmitter release and receptors, glucose metabolism, they modulate membrane channels (such as K^+ , Ca^{2+} and Cl^-) and stimulate neurogenesis [1170, 1171]; supplementation with blueberries increased hippocampal IGF1 and IGF1R levels and ERK activation in rats and these increases were significantly correlated with improved cognitive performance [1172] – this is of interest as the ERK pathway is activated by seizures and is thought to play a role in regulating neuronal excitability and synaptic plasticity [1173, 1174]. Activation of AKT has also been shown to promote neuronal survival [1175]. Impaired activation of these pathways may increase vulnerability of neurons to excitotoxic damage [1176]. On the other hand, although IGF1 is commonly cited as a neuroprotective agent [1177], it can increase neuronal excitability [1178], which could feasibly increase likelihood of a seizure.

One would expect the KD to down-regulate insulin/IGF1 signalling due to reduced glucose intake and the fact that increased responsiveness of the liver to insulin would lead to a decrease in gluconeogenesis, glycogenolysis and ketogenesis; these survival adaptations are necessary in order to reduce non-essential glucose utilisation and increase production of alternative energy sources (ketones), particularly for the brain. This is known as 'starvation diabetes' [1179] and very-low calorie diets and prolonged fasting have been shown to cause insulin resistance and diabetes in humans [1180]. In support of this, it has been shown that dietary fat intake is positively associated with serum IGF1 levels, whereas carbohydrate intake has a negative association [1181], and rodents subjected to calorie restriction have reduced serum IGF1 levels [1182] and increased activated AMPK, with subsequent down-regulation of AKT and mTOR activity [1183]. However, in humans, long-term calorie restriction was not associated with reduced serum IGF1 levels but did improve insulin sensitivity [1184]; in rat β cells, *Igf1r* expression was suppressed by both very high and very low glucose concentrations [1185]; KD-feeding in rats led to upregulation of brain *Igf1r* expression in rats, which did not occur with a carbohydrate-rich, calorie-restricted diet [411], and lead to increased AMPK signaling in the liver but not the brain [1186]; serum IGF1 has been shown to be decreased in children following the KD for three months, although levels did not decrease any further from this point onwards (follow-up of one year) [1187]. Increased insulin/IGF1 signalling may be a compensatory mechanism in response to chronic low glucose/insulin levels with the KD, which may involve PI3/AKT activation, SIRT1 and/or FOXO1: increased PI3/AKT activation and SIRT1 (which modulates transcription of some regulatory subunits of PI3K) have been implicated in the mechanisms of improved insulin sensitivity in skeletal muscle due to greater recruitment of GLUT4 to the cell membrane and thus greater glucose uptake [1188-1190]; calorie restriction-induced increased insulin sensitivity may also be due to activation of FOXO1, which leads to increased transcription of the insulin receptor and IRS2 [1191]; the increased fat intake

could also be responsible as increased phosphorylation of ERK1/2 is associated with elevated plasma NEFA levels [1192]. There may be tissue-specific effects on these signalling pathways induced by the KD, as has been shown with calorie restriction in rats [1193].

It may be the case that certain people respond well to the KD because their seizures are (at least partly) a result of impaired brain glucose utilisation and energy metabolism, which is overcome by high-fat, low-carbohydrate feeding. Impaired insulin sensitivity has been linked with neurodegeneration, in particular Alzheimer's disease, potentially due to its reduced glucose utilisation, reduced activation of AKT and thus increased neuronal death (AKT inhibits apoptosis by phosphorylating and thus inactivating GSK-3 β and BAD), increased oxidative stress, inflammation and hyper-phosphorylation of tau [1171, 1194, 1195]. Impaired insulin sensitivity is also associated with cognitive dysfunction, potentially due to reduced expression of acetylcholine transferase, an enzyme responsible for acetylcholine synthesis, a critical neurotransmitter in cognitive function, or due to increased release and decreased degradation of amyloid β -peptide metabolism with increased amyloid deposition [1196, 1197]. There are few reports of insulin resistance being associated with seizures [1198, 1199] but it is of interest that children with drug-resistant epilepsy have been shown to be more likely to be glucose intolerant, compared to those with epilepsy controlled by AEDs, independent of the AEDs being taken [1200]; the authors commented that the individuals with refractory epilepsy and abnormal oral glucose tolerance tests would be candidates for KD treatment. The most severe case of impaired brain glucose availability/metabolism linked with neurological disease (often with seizures) is GLUT1 deficiency secondary to *SLC2A1* mutations but, considering the neuroprotective role of insulin/IGF1 pathways, it is feasible that genetic variation influencing any component of these pathways could lead to brain-resistance to insulin/glucose signalling and potentially seizures. In this way, individuals harbouring variants that impair the action

of these pathways (and this is the reason for their seizures) would be expected to respond well to the KD. Alternatively, if the cause of seizures in some individuals is due to impaired brain glucose metabolism (for whatever reason) and the KD exerts an antiseizure effect via insulin/IGF1 pathways, individuals harbouring variants that impair the action of these pathways would not be expected to respond well to the KD. It may, of course, be the case that individuals with variants in these pathways have an impaired ability to put in place the neuroprotective mechanisms induced by seizures and so they are resistant to any type of antiepileptic treatment.

Variation in the triglyceride biosynthesis pathway was associated with favourable KD response in the exome sequencing analysis. Although not highlighted in analyses from the SNP chip, this process is closely related to insulin signalling and its downstream pathways. When activated by insulin, the transcription factor SREBP-1c stimulates lipogenesis (conversion of acetyl CoA, produced from the breakdown of excess carbohydrate, to fatty acids) and subsequently triglyceride synthesis (esterification of fatty acids with glycerol to form triglycerides). The effects of insulin on SREBP-1c are mediated by PI3-dependent pathways [1201]. Due to the high-fat content of the KD, free fatty acid levels inevitably increase and this is a sign of increased ketogenesis and use of ketone bodies as energy substrates. Too high levels of free fatty acids would lead to increased re-esterification of fatty acids and storage as triglycerides, once energy requirements have been met. Although high-fat diets have been found to stimulate lipogenesis [1202], due to the low-carbohydrate nature of the KD (which limits lipogenesis), together with the enhanced disposal of fatty acids via ketogenesis and mitochondrial β -oxidation, a KD with the appropriate amount of calories should not result in a net accumulation of triglycerides. Increased triglyceride levels are a hallmark feature of insulin resistance and depositing of triglycerides in tissues, such as skeletal muscle and liver, further exacerbates insulin

resistance. Deletion of certain genes present in the triglyceride biosynthesis pathway has been shown to protect against insulin resistance. For example, in *Elovl6*-knockout mice fed a high-fat, high-carbohydrate diet, insulin sensitivity was higher than wild types [1203]; this was thought to be due to recovery of the hepatic Irs2/Akt signalling pathway: hepatic Akt was still phosphorylated and this was accompanied by increased hepatic IRS2. Other genes, such as *ACSL1*, have not been shown to protect against insulin resistance in animal knockout models [1204]. It is possible that genetic variation that affects triglyceride biosynthesis favourably influences KD response due to putative effects on insulin sensitivity, but not every gene may have the same effect. It is unknown whether brain insulin sensitivity may be affected directly, but it is plausible that the systemic effects of the KD indirectly impact relevant pathways in the brain.

8.2 Study limitations

8.2.1 Classification of KD response

Clinic letters and dietetic notes were the principal source of information used to classify KD response. These were not created for research purposes and so do not always provide consistent data. The clinic letters used to mine information were written by different consultants, many of which were from different hospitals, who have different ways of writing clinic letters: some may give more information than others. The amount of detail provided even varies amongst those composed by the same clinician. These disadvantages must be balanced against the ease of obtaining such data; using notes already collected as part of clinical monitoring is less burdensome than asking participants to complete seizure diaries and record other information. This study was not a clinical trial and so it would have been more difficult to obtain ethical approval if such requirements were included in the protocol. Furthermore, the fact that the researcher was present in clinic appointments for many prospectively-recruited individuals meant that any information that was not

mentioned in clinic letters could still be used. Some individuals who followed the KD in the past and were recruited retrospectively, chose (or their parent/guardian(s) chose) to give details of their experience with the KD during recruitment, but such information was not systematic and could not be used to classify response at certain time points (it is highly unlikely that people remember exactly when seizure control was better or worse).

It should be noted that any non-significant results in genetic analyses may have been due to misclassification bias, if some participants were placed in the incorrect response group. Taking seizure-freedom (for example, at the 3-month point or sustained over a 2-year period) as the phenotype may have resulted in less misclassification bias, as there is less margin for error with regards seizure frequency when reporting such an extreme response. The problem with such an approach is that, due to the rarity of achieving seizure-freedom with the KD, a much larger sample size would have been needed to have confidence in any results obtained; in the current cohort, only nine individuals were seizure-free at the 3-month point and only two individuals maintained 100% seizure-freedom over a 2-year period. Furthermore, seizure-freedom whilst following the KD may be due to spontaneous remission, although one would encounter this problem when adopting any definition of response, due to the fluctuating nature of epilepsy. The genetic architecture of KD response is unknown and so it makes sense to start with a percentage seizure-reduction that is viewed as clinically useful in such a drug-resistant population; other phenotypes may subsequently be explored.

8.2.2 Compliance to treatment

Misclassification bias may also be present because some non-compliers were incorrectly classified as non-responders; this may have influenced genetic analyses. In clinical practice, the presence of ketones in urine and blood is often used as a sign of treatment compliance but these measures are subject to wide intra- and inter-individual variability. The

researcher tried to address this issue with the adapted MARS questionnaire but these data were not available for all participants and the questionnaire itself may not accurately reflect dietary compliance, as people aim to give socially-acceptable answers. The MARS measures intentional and unintentional non-adherence and there may have been inter-individual variability in interpretation of the question 'We alter the diet'; it may not have been clear whether or not this referred to changes prescribed by the dietitian. In this study, few changes were made in order to remain as close as possible to the validated versions of the questionnaire, but the MARS may have benefitted from further adaptation. Ideally, the adapted questionnaire should have been validated with more objective measures of compliance, such as weighed food records or doubly-labelled water. Accurate knowledge of food consumption may be obtained by patient surveillance and weighing of meals in future participants, as described in a previous case report¹⁸, but this may not be feasible on a large-scale out-patient basis.

8.2.3 Prospective versus retrospective recruitment

If all participants were recruited prospectively, completion of seizure diaries could have been stipulated as a requirement in Patient Information Sheets in order to improve the quality of KD response classification. Other tools could also have been used, such as quality of life and seizure severity questionnaires, which would have allowed more than one aspect to be taken into account when classifying KD response, although it is unknown whether this would have been viewed as acceptable by the Research Ethics Committee. As demonstrated by the phenotypic analyses in this work, other clinically relevant factors besides seizure frequency were lost to appraisal in this study because they were not systematically recorded or quantified. Due to limited clinic time, it may be a good idea to reserve a clinic room where patients could fill in such questionnaires whilst waiting for their appointments, with the researcher available to answer questions if needed. A balance

would have to be struck between obtaining as much data as possible and not over-burdening the patient or his/her parent/guardian(s).

Prospective recruitment would also have allowed better measures of compliance to be put in place. This is important as non-significant genetic results may be due to the fact that some non-compliers were incorrectly classified as non-responders. With prospective recruitment, the researcher could ensure that blood ketone levels are measured when patients attend clinics (this was done, where possible, but not for patients recruited retrospectively) or request recordings of daily ketone monitoring since the previous clinic visit. Completion of compliance questionnaires could be requested at various follow-up times from all participants, although this method does have its limitations (as noted in the previous section). Other methods, such as weighed food records or even short dietary recall interviews would provide further assurance that participants were actually following the KD. With a multi-centre study such as this, co-operation from colleagues at other centres would be required.

The benefit of retrospective recruitment is that the sample size can be more easily increased in a shorter space of time. Many individuals who previously followed the KD are still under follow-up with the same neurology team, or at least within the same centre.

The recruitment rate (whether prospective or retrospective) may have been greater if certain people at each site were involved from the very beginning. For example, dietitians, consultants or research assistants could have contributed to the protocol, giving them a sense of ownership.

8.2.4 Sample size

The size of the recruited cohort for this study was small, at least in terms of genetic analysis of a complex trait. This is the difficulty when the majority of cases are recruited by a single

person (despite putting systems in place for recruitment to occur elsewhere) and when the pool from which to recruit individuals in the first place is very limited - only several hundred people are thought to be following the KD during a year in the UK, and people who have followed the KD in the past are often lost to follow-up. The smaller the sample size, the less power there is to detect a significant genotypic-phenotypic association; although one can still examine the proportion of variants in responders and non-responders of the genes of most interest, rather than looking at p-values alone, other variants/genes/pathways with lesser effect sizes may also be relevant. The fact that some statistically significant results were obtained in this study, even with such a small sample size, is certainly encouraging. However, these results are largely dependent on which statistical test or definition of KD response is used. The more consistent the results and the larger the cohort, the more confident one can be that certain genes truly play a role in KD response. As it stands, caution is warranted when interpreting the reported results, given that this is a first report based on a relatively small sample.

8.2.5 Population structure

It is standard practice to adjust for population structure in genetic association studies. However, as shown in a study from the Wellcome Trust Case Consortium, population structure may only have a very small confounding effect: results from tests that adjusted for population structure were very similar to those from tests that did not adjust for population structure, and estimates of over-dispersion of the association statistics were close to 1 [573]. Individuals recruited for the KD study had a wide range of ethnicities, but the axes of variation explaining their genetic diversity was not correlated with the phenotype. λ_{GC} and quantile-quantile plots both showed that the least biased approach was a Fisher's test, which does not adjust for population stratification; this was used for the single variant tests using genotyping array and exome sequencing data, and the gene- and

pathway-based tests using genotyping array data. This is likely to do with the fact that the Fisher's test does not rely on an asymptotic distribution method: exact p-values are calculated based on the exact distribution of the test statistic (the dataset may, in fact, be unbalanced), rather than estimating p-values based on the assumption that the data conforms to a normal distribution (these approximations become more exact as the sample size nears infinity). With a larger sample size, population structure may become more of an issue. Circulating insulin levels and insulin sensitivity differs between ethnic populations – for example, African Americans tend to be hyperinsulinaemic and have an elevated insulin response and reduced insulin sensitivity compared to Caucasians [1205-1208]; this further strengthens the importance of examining the effect of population structure on KD response in future analyses and, if necessary, correcting for it.

8.2.6 Replicates and batch effect

Ideally, some individuals should have been genotyped or sequenced more than once.

Although stringent quality-control criteria were adopted and performance metrics did not differ between batches, high concordance would provide further reassurance that call and error rates did not differ between batches. Further funding would have to be sought in order to do this in the future.

Principal component analysis can be used to detect patterns in the data that reflect subgroups other than those defined by ethnicity, such as batch effect, and included as covariates if necessary.

8.3 Future work

8.3.1 Continuing recruitment and further analyses

If possible, recruitment should be extended in order to obtain as large a sample size as possible – preferably ≥ 1000 participants. Further international collaboration and dedicated

research assistants, or at least somebody responsible for recruitment at each site, would greatly aid this effort.

Ideally, recruitment would be prospective in order to allow systematic collection of data so that compliance could be more closely monitored, seizure frequency could be more accurately quantified and other components besides seizure frequency could be used when classifying KD response. This would require submission of an amendment to the Research Ethics Committee. If a large enough cohort was recruited, separate analyses could be conducted to determine whether the mechanisms that lead to, for example, increased alertness with the KD are different to those that lead to reduced seizure frequency.

KD response data should be collected for as many time points as possible, as has been done with this cohort. This would allow future analyses to be conducted with response classified at specific time points, or accounting for longitudinal response. There may be a genetic reason why some individuals do not respond well at all to the KD, others respond well to the KD in the short-term but then relapse, whereas others have a sustained response.

With a larger sample size, other approaches to analysing the genetic data could be taken. For example, it may be possible to look for the genetic cause of epilepsy in individuals for whom this is unknown (by identifying rare variants of putative functional consequence in genes that are known to be associated with epilepsy) and investigate whether this is correlated with KD response. This could feasibly be done with whole exome sequencing or with a genotyping array with additional epilepsy-specific markers. Association analyses could also be conducted with individuals with specific (non-genetic) causes of epilepsy; in this way, each sub-cohort, each with a specific diagnosis, would be more homogeneous than the present cohort, thus reducing potential 'noise'. A large sample size would be needed for such analyses because individuals that commence dietary treatment for

epilepsy are very heterogeneous; one would require as many individuals as possible in each group, most or all with a similar response to the KD, in order to increase confidence that the genetic cause of their epilepsy is also correlated with KD response, or that the mechanism for KD response is distinct in individuals with certain epilepsy aetiologies.

Other programmes could also be used to conduct gene- and pathway-based analyses, both with genotyping array and exome sequencing data, and overlapping results prioritised.

8.3.2 Gene expression analyses

In addition to overcoming the sample size limitation of the gene variant analyses, investigation into what gene expression changes occur in humans following the KD (*Research question 3*) would be of great interest. Comparing gene expression levels in individuals before they start the KD, when they have followed the KD for a certain length of time, and potentially soon after being weaned from the diet, could answer two questions:

- i) what changes in gene expression does the KD induce in humans? Do up- or down-regulated genes have a particular function in common or belong to a particular pathway? This may or may not be the same as changes induced by starvation/calorie restriction alone, or by KD-feeding in animal studies.
- ii) do KD responders have different gene expression changes compared to non-responders? This could be examined both separately and alongside results from the SNP chip and/or exome sequencing. It may be that certain gene expression changes can be linked with specific gene variants.

If RNA were to be collected at various time points, it may be determined whether certain gene expression changes are more correlated with short- or long-term response. If RNA was also collected after discontinuation of the KD, it may be determined how long KD-induced changes in gene expression last. It may be the case that those who discontinue

treatment following a long period of sustained response (normally 2+ years) are more or less likely to maintain the benefits obtained from the KD, depending on these long-term expression changes. RNA from brain would be ideal, although this is relatively inaccessible. Gene expression in peripheral blood lymphocytes has been shown to moderately correlate with gene expression in the brain [1209]. Therefore, except for genes with tissue-specific expression, peripheral blood gene expression data may be used as a surrogate for CNS gene expression. It would be desirable to develop blood-based biomarkers for KD response due to the relative non-invasive nature of whole blood collection, or at least determine whether genes with certain functions are more likely to be differentially regulated in responders and non-responders.

8.3.3 Follow-up studies

If certain genes were highlighted as potentially relevant in higher-powered analyses with a larger sample size and, potentially, in another replication cohort, targeted genotyping or sequencing of the regions of interest could then be undertaken with future samples, rather than having to obtain whole genome or whole exome data. Gene knockout functional assays could help ascertain the phenotypic consequences of gene ablation, including general consequences, such as whether this leads to apoptosis or influences cell proliferation, or whether more specific features of the gene/protein, such as secreting a particular hormone, remain intact or not. It would be of interest to see whether organ-specific gene knockout (for example, in the brain versus the liver or pancreas) had different effects. Functional studies could also help determine the effects of certain variants on gene/protein function.

These approaches, however, are problematic when dealing with a complex trait: firstly, because there may be many genes of interest, which would increase cost of future analyses and, secondly, because investigation of individual genes does not account for the fact that

the KD is likely influenced by metabolic or cell signalling networks. Results have shown that one single gene may have various functions and influence various pathways that may feasibly be relevant to KD response. Genes that appear to have the largest effect size on the phenotypic trait should be identified for further analysis, whilst bearing in mind that no single gene is likely to fully explain the whole story.

8.3.4 Clinical translation

If one or more genes were found to have a large effect size on KD response, either from gene variant or expression analyses, results from targeted genotyping or sequencing (conducted in a clinically-accredited laboratory) could be used as an additional factor when deciding whether to commence KD treatment for an individual at a certain point in time. If abnormal function of these genes, or of the pathways within which they function, could be detected by specific biomarkers, this would greatly facilitate this process. This is unlikely, because no parameter that is routinely measured before or during KD treatment has been consistently shown to correlate with response. In light of the results obtained, I would be interested in assessing whether glucose tolerance or insulin sensitivity could be used as a biomarker to predict KD response.

8.4 Final words

Four themes have emerged from the various genetic analyses conducted, including variant-, gene- and pathway-based analyses: cell cycle progression and apoptosis, glucose homeostasis, neurological processes and triglyceride biosynthesis. There was no evidence that common variation has a greater or lesser effect than rare variation on KD response, as these themes were identified from all analyses. It is of interest that the first three themes highlighted are also linked with the candidate genes *KCNJ11* and *BAD*, although variants in other genes appeared to have a greater differential distribution amongst responders and

non-responders. These four themes can all be traced back in some way to the insulin signalling cascade. However, the implications of these pathways are wide (for example, insulin regulates the expression of over 100 genes [1210] and there are over 1000 combinatorial possibilities only in the first steps of the insulin/IGF1 pathways [1211]), even if one just concentrates on the effects on the brain: neuronal survival, brain glucose metabolism, neurotransmitter release and so on. As such, genetic variation in many genes downstream of insulin/IGF1, PI3/AKT/mTOR and MAPK/ERK pathways may influence these diverse effects. Impaired function of certain components of a pathway may be compensated for by others, which hinders the use of loss-of-function strategies.

One may claim that this work has done little to enhance understanding of the mechanisms underlying how the KD restores neuronal excitability homeostasis in some individuals but not others. Despite the study limitations, I would argue against this; the results obtained provide evidence that KD response is influenced by a web of factors, many of which are inter-related. In subsets of individuals, unfavourable KD response may be explained by a specific variant that, for example, influences glutamate or glucose metabolism; in other individuals, there may be many biologically feasible explanations for their favourable or unfavourable KD response but, when taken in isolation, no theory solely accounts for this variability. The major disadvantage of this is that, at this stage, these results have little clinical significance. Although it is likely that KD response is influenced by genetic variation, the mechanisms are too complex to be narrowed down to a single gene and are likely to involve multiple pathways. Certain genes may be more relevant than others and this study has highlighted several genes for further investigation. A larger sample size and additional studies are certainly needed to increase confidence in the importance of these genes.

9 References

1. Fisher, R.S., et al., *Epileptic Seizures and Epilepsy: Definitions Proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE)*. *Epilepsia*, 2005. **46**(4): p. 470-472.
2. Berg, A.T., et al., *Revised terminology and concepts for organization of seizures and epilepsies: report of the ILAE Commission on Classification and Terminology, 2005-2009*. *Epilepsia*, 2010. **51**(4): p. 676-85.
3. Ngugi, A.K., et al., *Estimation of the burden of active and life-time epilepsy: a meta-analytic approach*. *Epilepsia*, 2010. **51**(5): p. 883-90.
4. Ngugi, A.K., et al., *Incidence of epilepsy: a systematic review and meta-analysis*. *Neurology*, 2011. **77**(10): p. 1005-12.
5. Guerrini, R., *Epilepsy in children*. *The Lancet*, 2006. **367**(9509): p. 499-524.
6. Kwan, P., Schachter, S.C., and Brodie, M.J., *Drug-resistant epilepsy*. *N Engl J Med*, 2011. **365**: p. 919-26.
7. Kwan, P., et al., *Definition of drug resistant epilepsy: consensus proposal by the ad hoc Task Force of the ILAE Commission on Therapeutic Strategies*. *Epilepsia*, 2010. **51**(6): p. 1069-77.
8. Gustavsson, A., et al., *Cost of disorders of the brain in Europe 2010*. *European neuropsychopharmacology : the journal of the European College of Neuropsychopharmacology*, 2011. **21**(10): p. 718-779.
9. Begley, C.E., et al., *Early treatment cost in epilepsy and how it varies with seizure type and frequency*. *Epilepsy Research*, 2001. **47**(3): p. 205-215.
10. van Hout, B., et al., *Relationship between seizure frequency and costs and quality of life of outpatients with partial epilepsy in France, Germany, and the United Kingdom*. *Epilepsia*, 1997. **38**(11): p. 1221-6.
11. Kobau, R., et al., *Satisfaction with life domains in people with epilepsy*. *Epilepsy Behav*, 2012. **25**(4): p. 546-551.
12. Taylor, R.S., et al., *Predictors of health-related quality of life and costs in adults with epilepsy: A systematic review*. *Epilepsia*, 2011. **52**(12): p. 2168-2180.
13. Bautista, R.E. and Glen, E., *Seizure severity is associated with quality of life independent of seizure frequency*. *Epilepsy Behav*, 2009. **16**(2): p. 325-329.
14. Nadkarni, J., Jain, A., and Dwivedi, R., *Quality of life in children with epilepsy*. *Ann Indian Acad Neurol*, 2011. **14**(4): p. 279-82.
15. Thomas, S.V., et al., *Frequent seizures and polytherapy can impair quality of life in persons with epilepsy*. *Neurol India*, 2005. **53**(1): p. 46-50.
16. Sabaz, M., et al., *Validation of a new quality of life measure for children with epilepsy*. *Epilepsia*, 2000. **41**(6): p. 765-74.
17. Devinsky, O., et al., *Risk factors for poor health-related quality of life in adolescents with epilepsy*. *Epilepsia*, 1999. **40**(12): p. 1715-20.
18. Berg, A.T., et al., *Age at onset of epilepsy, pharmacoresistance, and cognitive outcomes: A prospective cohort study*. *Neurology*, 2012. **79**(13): p. 1384-91.
19. Nickels, K.C., Grossardt, B.R., and Wirrell, E.C., *Epilepsy-related mortality is low in children: A 30-year population-based study in Olmsted County, MN*. *Epilepsia*, 2012. **53**: p. 2164-71.
20. National Clinical Guideline Centre, *The Epilepsies: The diagnosis and management of the epilepsies in adults and children in primary and secondary care*, 2012, Commissioned by the National Institute for Health and Clinical Excellence: National Clinical Guideline Centre, London.

21. NICE, *The epilepsies: the diagnosis and management of the epilepsies in adults and children in primary and secondary care (update)*. 2012.
22. Walia, K.S., et al., *Side effects of antiepileptics--a review*. Pain Pract, 2004. **4**(3): p. 194-203.
23. Best, J.L. and Acheson, J.F., *The natural history of Vigabatrin associated visual field defects in patients electing to continue their medication*. Eye (Lond), 2005. **19**(1): p. 41-4.
24. Holliday, S.L. and R.L. Brey, *Patient page. Memory problems after epilepsy surgery*. Neurology, 2003. **60**(6): p. E3-5.
25. Manji, H., *Epilepsy surgery, visual fields, and driving: a study of the visual field criteria for driving in patients after temporal lobe epilepsy surgery with a comparison of goldmann and esterman perimetry*. Am J Ophthalmol, 2000. **129**(5): p. 704.
26. Kwan, P. and Sperling, M.R., *Refractory seizures: try additional antiepileptic drugs (after two have failed) or go directly to early surgery evaluation?* Epilepsia, 2009. **50 Suppl 8**: p. 57-62.
27. Klinkenberg, S., et al., *Vagus nerve stimulation in children with intractable epilepsy: a randomized controlled trial*. Dev Med Child Neurol, 2012. **54**(9): p. 855-61.
28. Klinkenberg, S., et al., *Behavioural and cognitive effects during vagus nerve stimulation in children with intractable epilepsy - a randomized controlled trial*. Eur J Paediatr Neurol, 2013. **17**(1): p. 82-90.
29. Ben-Menachem, E., *Vagus nerve stimulation, side effects, and long-term safety*. J Clin Neurophysiol, 2001. **18**(5): p. 415-8.
30. Katchanov, J. and Birbeck, G.L., *Epilepsy care guidelines for low- and middle-income countries: From WHO mental health GAP to national programs*. BMC Med, 2012. **10**(1): p. 107.
31. Ehrhardt, S. and Meyer, C.G., *Transfer of evidence-based medical guidelines to low- and middle-income countries*. Trop Med Int Health, 2012. **17**(2): p. 144-6.
32. WHO, *mhGAP Intervention Guide for mental, neurological and substance use disorders in non-specialized health settings*. 2011.
33. Guelpa, G. and Marie, A., *La lutte contre l'épilepsie par la désintoxication et par la reéducation alimentaire [The fight against epilepsy by detoxification and by food reduction]*. Rev Ther Med Chir (Paris), 1911. **78**: p. 8-13.
34. Wilder, R.M., *The effect on ketonemia on the course of epilepsy*. Bull Mayo Clin, 1921. **2**: p. 307-308.
35. Keith, D.A., *Side effects of diphenylhydantoin: a review*. J Oral Surg, 1978. **36**(3): p. 206-9.
36. Huttenlocher, P.R., *Ketonemia and seizures: metabolic and anticonvulsant effects of two ketogenic diets in childhood epilepsy*. Pediatr Res, 1976. **540**: p. 536-540.
37. Huttenlocher, P.R., Wilbourn, A.J., and Signore, J.M., *Medium-chain triglycerides as a therapy for intractable childhood epilepsy*. Neurology, 1971. **21**: p. 1097-1103.
38. Schwartz, R.H., et al., *Ketogenic diets in the treatment of epilepsy: short term clinical effects*. Dev Med Child Neurol, 1989. **31**: p. 145-151.
39. Cederbaum, S.D., et al., *Sensitivity to carbohydrate in a patient with familial intermittent lactic acidosis and pyruvate dehydrogenase deficiency*. Pediatr Res, 1976. **10**(8): p. 713-20.
40. Wexler, I.D., et al., *Outcome of pyruvate dehydrogenase deficiency treated with ketogenic diets. Studies in patients with identical mutations*. Neurology, 1997. **49**(6): p. 1655-61.

41. De Vivo, D.C., et al., *Defective glucose transport across the blood-brain barrier as a cause of persistent hypoglycorrhachia, seizures, and developmental delay*. N Eng J Med, 1991. **325**(10): p. 703-9.
42. Wheless, J.W., *History of the ketogenic diet*. Epilepsia, 2008. **49** (Suppl 8): p. 3-5.
43. Kossoff, E.H. and Dorward, J.L., *The modified Atkins diet*. Epilepsia, 2008. **49 Suppl 8**: p. 37-41.
44. Groesbeck, D.K., Bluml, R.M., and Kossoff, E.H., *Long-term use of the ketogenic diet in the treatment of epilepsy*. Dev Med Child Neurol, 2006. **48**(12): p. 978-81.
45. Lefevre, F. and N. Aronson, *Ketogenic diet for the treatment of refractory epilepsy in children: A systematic review of efficacy*. Pediatrics, 2000. **105**(4): p. E46.
46. Keene, D.L., *A systematic review of the use of the ketogenic diet in childhood epilepsy*. Pediatr Neurol, 2006. **35**(1): p. 1-5.
47. Henderson, C.B., et al., *Efficacy of the ketogenic diet as a treatment option for epilepsy: meta-analysis*. J Child Neurol, 2006. **21**(3): p. 193-8.
48. Sharma, S., et al., *Use of the modified Atkins diet for treatment of refractory childhood epilepsy: a randomized controlled trial*. Epilepsia, 2013. **54**(3): p. 481-6.
49. Freeman, J.M., et al., *A blinded, crossover study of the efficacy of the ketogenic diet*. Epilepsia, 2009. **50**: p. 322-325.
50. Kossoff, E.H., et al., *A randomized, crossover comparison of daily carbohydrate limits using the modified Atkins diet*. Epilepsy Behav, 2007. **10**: p. 432-6.
51. Seo, J.H., et al., *Efficacy and tolerability of the ketogenic diet according to lipid:nonlipid ratios - comparison of 3:1 with 4:1 diet*. Epilepsia, 2007. **48**: p. 801-5.
52. Bergqvist, A.G.C., et al., *Fasting versus gradual initiation of the ketogenic diet: a prospective, randomized clinical trial of efficacy*. Epilepsia, 2005. **46**: p. 1810-9.
53. Fraser, D.D., et al., *from children on the ketogenic diet Elevated polyunsaturated fatty acids in blood serum obtained from children on the ketogenic diet*. Neurology, 2003. **60**: p. 1026-1029.
54. Villeneuve, N., et al., *The ketogenic diet improves recently worsened focal epilepsy*. Dev Med Child Neurol, 2009. **51**: p. 276-81.
55. Kossoff, E.H., et al., *Prospective study of the modified atkins diet in combination with a ketogenic liquid supplement during the initial month*. J Child Neurol, 2011. **26**: p. 147-51.
56. Ross, D.L., et al., *Early biochemical and EEG correlates of the ketogenic diet in children with atypical absence epilepsy*. Pediatr Neurol, 1985. **1**: p. 104-108.
57. Kossoff, E.H., et al., *A case-control evaluation of the ketogenic diet versus ACTH for new-onset infantile spasms*. Epilepsia, 2008. **49**: p. 1504-9.
58. Caraballo, R., et al., *Dieta cetogénica: Eficacia y tolerabilidad en epilepsia refractaria en pediatría [Ketogenic diet: Efficacy and Tolerability in Childhood intractable epilepsies]*. Rev Neurologia, 1998. **26**: p. 61-64.
59. Edelstein, S.F. and Chisholm, M., *Management of Intractable Childhood Seizures Using the Non-MCT Oil Ketogenic Diet in 20 Patients*. J Am Diet Assoc, 1996. **96**: p. 1181-1182.
60. Mak, S.C., Chi, C.S., and Wan, C. J., *Clinical experience of ketogenic diet on children with refractory epilepsy*. Acta Paediatr Taiwan, 1999. **40**(2): p. 97-100.
61. Doolittle, G.J., *Report of results from use of ketogenic diet and ketogenic diet with water restriction in a series of epileptics*. Psychiatric Quarterly, 1931. **5**: p. 224-252.
62. Weeks, D.F., et al., *Observations on fasting and diets in the treatment of epilepsy*. J Metab Res, 1923. **3**: p. 317-364.
63. Bastible, C., *The ketogenic treatment of epilepsy*. Ir J Med Sci 1931. **2**: p. 506-520.
64. Swink, T.D., Vining, E.P., and Freeman, J.M., *The ketogenic diet: 1997*. Adv Pediatr, 1997. **44**: p. 297-329.

65. Freeman, J.M., et al., *The Efficacy of the Ketogenic Diet — 1998 : A Prospective Evaluation of Intervention in 150 Children*. Pediatrics, 1998. **102**: p. 1358-1363.
66. Vining, E.P., et al., *A multicenter study of the efficacy of the ketogenic diet*. Arch Neurol, 1998. **55**: p. 1433-1437.
67. Cusmai, R., et al., *La dieta chetogena nelle epilessie farmacoresistenti [Ketogenic diet in pharmacoresistant epilepsies]*. Bolletino Lega Italiana Contro l'Epilessia, 1999. **106**(7): p. 365-367.
68. Gilbert, D.L., Pyzik, P.L., and Freeman, J.M., *The Ketogenic Diet : Seizure Control Correlates Better With Serum & beta-Hydroxybutyrate Than With Urine Ketones*. J Child Neurol, 2000. **15**: p. 787-790.
69. Panico, L.R., et al., *Dieta cetogénica en la epilepsia refractaria infantil: respuesta electroclínica, complicaciones y efectos secundarios [The ketogenic diet in infantile refractory epilepsy: electroclinical response, complications and secondary effects]*. Rev Neurologia, 2000. **31**(3): p. 212-220.
70. Panico, L., et al., *Evolución electroencefalográfica de un grupo de pacientes en dieta cetogénica [The electroencephalographic evolution of a group of patients on a ketogenic diet]*. Rev Neurologia, 2000. **30**: p. 8-15.
71. Kankirawatana, P., et al., *Ketogenic diet: an alternative treatment for refractory epilepsy in children*. J Med Assoc Thai, 2001. **84**(7): p. 1027-1032.
72. Maydell, B.V., et al., *Efficacy of the ketogenic diet in focal versus generalized seizures*. Pediatr Neurol, 2001. **25**: p. 208-212.
73. Coppola, G., et al., *The ketogenic diet in children , adolescents and young adults with refractory epilepsy : an Italian multicentric experience*. Epilepsy Res, 2002. **48**: p. 221-227.
74. Dimario, F.J. and Holland, J., *The Ketogenic Diet: A Review of the Experience at Connecticut Children ' s Medical Center*. Pediatr Neurol, 2002. **26**: p. 288-292.
75. Kossoff, E.H., et al., *Efficacy of the Ketogenic Diet for Infantile Spasms*. Pediatrics, 2002. **109**: p. 780-783.
76. Mady, M.A., et al., *The ketogenic diet: adolescents can do it, too*. Epilepsia, 2003. **44**: p. 847-851.
77. Kim, D.W., et al., *Benefits of the nonfasting ketogenic diet compared with the initial fasting ketogenic diet*. Pediatrics, 2004. **114**: p. 1627-30.
78. Kossoff, E.H., et al., *The impact of early versus late anticonvulsant reduction after ketogenic diet initiation*. Epilepsy Behav, 2004. **5**: p. 499-502.
79. Caraballo, R.H., et al., *Ketogenic diet in patients with Dravet syndrome*. Epilepsia, 2005. **46**: p. 1539-44.
80. Hosain, S.A., La Vega-Talbott, M., and Solomon, G.E., *Ketogenic Diet in Pediatric Epilepsy Patients With Gastrostomy Feeding*. Pediatric Neurology, 2005. **32**: p. 81-83.
81. Kang, C.H., et al., *Efficacy and safety of the ketogenic diet for intractable childhood epilepsy: Korean multicentric experience*. Epilepsia, 2005. **46**: p. 272-279.
82. Kossoff, E.H., et al., *Tuberous sclerosis complex and the ketogenic diet*. Epilepsia, 2005. **46**: p. 1684-1686.
83. Lyczkowski, D.A., et al., *Safety and tolerability of the ketogenic diet in pediatric epilepsy: effects of valproate combination therapy*. Epilepsia, 2005. **46**: p. 1533-1538.
84. Mackay, M.T., et al., *The ketogenic diet in refractory childhood epilepsy*. J Paediatr Child Health, 2005. **41**: p. 353-357.
85. Rubenstein, J.E., et al., *Experience in the Use of the Ketogenic Diet as Early Therapy*. J Child Neurol, 2005. **20**: p. 31-34.

86. Eun, S., Kang, H., and Kim, D., *Ketogenic diet for treatment of infantile spasms*. Brain Dev, 2006. **28**: p. 566-571.
87. Peterson, S.J., et al., *Changes in growth and seizure reduction in children on the ketogenic diet as a treatment for intractable epilepsy*. J Am Diet Assoc, 2005. **105**: p. 718-725.
88. Bergqvist, A.G., et al., *Predictive power of first morning glucose and the ketogenic diet*. Neuropediatrics, 2007. **38**(4): p. 193-196.
89. Dahlin, M., et al., *Plasma phospholipid fatty acids are influenced by a ketogenic diet enriched with n-3 fatty acids in children with epilepsy*. Epilepsy Res, 2007. **73**: p. 199-207.
90. Freitas, A., et al., *Ketogenic diet for the treatment of refractory epilepsy: a 10 year experience in children*. Arq Neuropsiquiatr, 2007. **65**: p. 381-384.
91. Hallböök, T., et al., *Effects of ketogenic diet on epileptiform activity in children with therapy resistant epilepsy*. Epilepsy Res, 2007. **77**: p. 134-140.
92. Hallbook, T., Lundgren, J., and Ingmar, R., *Ketogenic Diet Improves Sleep Quality in Children with Therapy-resistant Epilepsy*. Epilepsia, 2007. **48**: p. 59-65.
93. Hamdy, R., et al., *Lack of influence of body mass index on the efficacy of the ketogenic diet*. J Child Neurol, 2007. **22**: p. 1167-1171.
94. Seo, J.H., et al., *Efficacy and tolerability of the ketogenic diet according to lipid:nonlipid ratios--comparison of 3:1 with 4:1 diet*. Epilepsia, 2007. **48**: p. 801-805.
95. Stainman, R.S., et al., *Decreased relative efficacy of the ketogenic diet for children with surgically approachable epilepsy*. Seizure, 2007. **16**: p. 615-9.
96. Jung, D.E., Kang, H., and Kim, H., *Long-term outcome of the ketogenic diet for intractable childhood epilepsy with focal malformation of cortical development*. Pediatrics, 2008. **122**: p. e330-e333.
97. Kossoff, E., et al., *When do seizures usually improve with the ketogenic diet?* Epilepsia, 2008. **49**: p. 329-333.
98. Neal, E.G., et al., *A randomized trial of classical and medium-chain triglyceride ketogenic diets in the treatment of childhood epilepsy*. Epilepsia, 2009. **50**: p. 1109-17.
99. Remahl, S., Dahlin, M.G., and Amark, P.E., *Influence of the ketogenic diet on 24-hour electroencephalogram in children with epilepsy*. Pediatr Neurol, 2008. **38**: p. 38-43.
100. Mosek, A., et al., *Ketogenic diet treatment in adults with refractory epilepsy: a prospective pilot study*. Seizure, 2009. **18**: p. 30-33.
101. Porta, N., et al., *Comparison of seizure reduction and serum fatty acid levels after receiving the ketogenic and modified Atkins diet*. Seizure, 2009. **18**: p. 359-64.
102. Sharma, S., et al., *Seizure control and biochemical profile on the ketogenic diet in young children with refractory epilepsy--Indian experience*. Seizure, 2009. **18**: p. 446-449.
103. Beniczky, S., et al., *Effectiveness of the ketogenic diet in a broad range of seizure types and EEG features for severe childhood epilepsies*. Acta Neurol Scand, 2010. **121**: p. 58-62.
104. Coppola, G., et al., *Original article Ketogenic diet for the treatment of catastrophic epileptic encephalopathies in childhood*. Eur J Paediatr Neurol, 2010. **14**: p. 229-234.
105. Hong, A., Turner, Z., and Hamdy, R., *Infantile spasms treated with the ketogenic diet: prospective single-center experience in 104 consecutive infants*. Epilepsia, 2010. **51**: p. 1403-1407.

106. Klein, P., et al., *Ketogenic diet treatment in adults with refractory epilepsy*. *Epilepsy Behav*, 2010. **19**: p. 575-579.
107. Kessler, S.K., et al., *Early EEG improvement after ketogenic diet initiation*. *Epilepsy Res*, 2011. **94**: p. 94-101.
108. Nababout, R., et al., *Ketogenic diet also benefits Dravet syndrome patients receiving stiripentol: a prospective pilot study*. *Epilepsia*, 2011. **52**: p. e54-e57.
109. Numis, A.L., et al., *The relationship of ketosis and growth to the efficacy of the ketogenic diet in infantile spasms*. *Epilepsy Res*, 2011. **96**: p. 172-5.
110. Raju, K.N., et al., *Efficacy of 4:1 (classic) versus 2.5:1 ketogenic ratio diets in refractory epilepsy in young children: a randomized open labeled study*. *Epilepsy Res*, 2011. **96**(1-2): p. 96-100.
111. Rosen, B., et al., *Cerebral changes after injury to the median nerve: a long-term follow up*. *J Plast Surg Hand Surg*, 2012. **46**(2): p. 106-12.
112. Thammongkol, S., et al., *Efficacy of the ketogenic diet: Which epilepsies respond?* *Epilepsia*, 2012. **53**: p. e55-e59.
113. Lemmon, M.E., et al., *Efficacy of the ketogenic diet in Lennox-Gastaut syndrome: a retrospective review of one institution's experience and summary of the literature*. *Dev Med Child Neurol*, 2012. **54**(5): p. 464-8.
114. Suo, C., et al., *Efficacy and safety of the ketogenic diet in Chinese children*. *Seizure*, 2013. **22**(3): p. 174-8.
115. Pires, M.E., et al., *Ketogenic diet for infantile spasms refractory to first-line treatments: An open prospective study*. *Epilepsy Research*, 2013. **105**(1-2): p. 189-94.
116. Li, B., et al., *Effects of ketogenic diet on the clinical and electroencephalographic features of children with drug therapy-resistant epilepsy*. *Exp Ther Med*, 2013. **5**(2): p. 611-615.
117. Moreno-Villares, J.M., et al., *La vuelta a la dieta cetogénica. ¿Qué papel desempeña en el tratamiento de las convulsiones infantiles refractarias? [The return to the ketogenic diet. What role does it play in the treatment of refractory seizures of infancy?]* *Rev Neurol*, 2001. **32**: p. 1115-1119.
118. Kossoff, E.H., et al., *A modified Atkins diet is effective for the treatment of intractable pediatric epilepsy*. *Epilepsia*, 2006. **47**: p. 421-424.
119. Kang, H., et al., *Use of a modified Atkins diet in intractable childhood epilepsy*. *Epilepsia*, 2007. **48**: p. 182-186.
120. Carrette, E., et al., *A pilot trial with modified Atkins' diet in adult patients with refractory epilepsy*. *Clin Neurol Neurosurg*, 2008. **110**: p. 797-803.
121. Kossoff, E.H., et al., *A prospective study of the modified Atkins diet for intractable epilepsy in adults*. *Epilepsia*, 2008. **49**: p. 316-9.
122. Weber, S., et al., *Modified Atkins diet to children and adolescents with medical intractable epilepsy*. *Seizure*, 2009. **18**: p. 237-40.
123. Kossoff, E.H., Borsage, J.L., and Comi, A.M., *A pilot study of the modified Atkins diet for Sturge-Weber syndrome*. *Epilepsy Res*, 2010. **92**: p. 240-3.
124. Tonekaboni, S.H., et al., *Efficacy of the Atkins diet as therapy for intractable epilepsy in children*. *Arch Iran Med*, 2010. **13**: p. 492-7.
125. Miranda, M.J., et al., *Danish study of a modified Atkins diet for medically intractable epilepsy in children: can we achieve the same results as with the classical ketogenic diet?* *Seizure*, 2011. **20**: p. 151-5.
126. Smith, M., et al., *Efficacy and tolerability of the modified Atkins diet in adults with pharmaco-resistant epilepsy: a prospective observational study*. *Epilepsia*, 2011. **52**: p. 775-80.

127. Chen, W. and Kossoff, E.H., *Long-term follow-up of children treated with the Modified Atkins Diet*. J Child Neurol, 2012. **27**: p. 754-758.
128. Muzykewicz, D.A., et al., *Efficacy, safety, and tolerability of the low glycemic index treatment in pediatric epilepsy*. Epilepsia, 2009. **50**: p. 1118-26.
129. Pfeifer, H.H., Lyczkowski, D.A., and Thiele, E.A., *Low glycemic index treatment: implementation and new insights into efficacy*. Epilepsia, 2008. **49 Suppl 8**: p. 42-5.
130. Larson, A.M., Pfeifer, H.H., and Thiele, E.A., *Low glycemic index treatment for epilepsy in tuberous sclerosis complex*. Epilepsy Res, 2012. **99**: p. 180-2.
131. Lambrechts, D.A., et al., *The ketogenic diet as a treatment option in adults with chronic refractory epilepsy: Efficacy and tolerability in clinical practice*. Epilepsy Behav, 2012. **23**: p. 310-314.
132. van Delft, R., et al., *Blood beta-hydroxybutyrate correlates better with seizure reduction due to ketogenic diet than do ketones in the urine*. Seizure, 2010. **19**: p. 36-9.
133. Hemingway, C., Freeman, J., and Pillas, D., *The ketogenic diet: a 3-to 6-year follow-up of 150 children enrolled prospectively*. Pediatrics, 2001. **108**: p. 898-905.
134. Martinez, C.C., Pyzik, P.L., and Kossoff, E.H., *Discontinuing the ketogenic diet in seizure-free children: recurrence and risk factors*. Epilepsia, 2007. **48**: p. 187-90.
135. Caraballo, R.H., *Nonpharmacologic treatments of Dravet syndrome: focus on the ketogenic diet*. Epilepsia, 2011. **52 Suppl 2**: p. 79-82.
136. Kang, H.C., et al., *Comparison of short- versus long-term ketogenic diet for intractable infantile spasms*. Epilepsia, 2011. **52**: p. 781-787.
137. Patel, A., et al., *Long-term outcomes of children treated with the ketogenic diet in the past*. Epilepsia, 2010. **51**: p. 1277-82.
138. Marsh, E.B., et al., *The outcome of children with intractable seizures: a 3- to 6-year follow-up of 67 children who remained on the ketogenic diet less than one year*. Epilepsia, 2006. **47**: p. 425-30.
139. Sirven, J., et al., *The ketogenic diet for intractable epilepsy in adults: preliminary results*. Epilepsia, 1999. **40**: p. 1721-6.
140. Yancy, W.S., Jr., et al., *Effects of two weight-loss diets on health-related quality of life*. Qual Life Res, 2009. **18**(3): p. 281-9.
141. Halyburton, A.K., et al., *Low- and high-carbohydrate weight-loss diets have similar effects on mood but not cognitive performance*. Am J Clin Nutr, 2007. **86**(3): p. 580-587.
142. McClernon, F.J., et al., *The Effects of a Low-Carbohydrate Ketogenic Diet and a Low-Fat Diet on Mood, Hunger, and Other Self-Reported Symptoms*. Obesity, 2007. **15**(1): p. 182-182.
143. Nordli, D.R., et al., *Experience with the ketogenic diet in infants*. Pediatrics, 2001. **108**: p. 129-133.
144. Peterman, M., *The ketogenic diet in epilepsy*. JAMA, 1925. **84**: p. 1979-1983.
145. Kinsman, S.L., et al., *Efficacy of the ketogenic diet for intractable seizure disorders: review of 58 cases*. Epilepsia, 1992. **33**: p. 1132-6.
146. Pulsifer, M.B., et al., *Effects of ketogenic diet on development and behavior: preliminary report of a prospective study*. Dev Med Child Neurol, 2001. **43**(5): p. 301-306.
147. Farasat, S., et al., *The importance of parental expectations of cognitive improvement for their children with epilepsy prior to starting the ketogenic diet*. Epilepsy Behav, 2006. **8**(2): p. 406-410.
148. Lambrechts, D.A., et al., *Ketogenic diet effects on cognition, mood, and psychosocial adjustment in children*. Acta Neurol Scand, 2013. **127**(2): p. 103-8.

149. Weber, S., et al., *Modified Atkins diet to children and adolescents with medical intractable epilepsy*. *Seizure*, 2009. **18**(4): p. 237-40.
150. Lightstone, L., et al., *Reasons for failure of the ketogenic diet*. *J Neurosci Nurs*, 2001. **33**(6): p. 292-295.
151. Kossoff, E.H., et al., *Optimal clinical management of children receiving the ketogenic diet: recommendations of the International Ketogenic Diet Study Group*. *Epilepsia*, 2009. **50**: p. 304-17.
152. Henderson, C.B., et al., *Efficacy of the ketogenic diet as a treatment option for epilepsy: meta-analysis*. *J Child Neurol*, 2006. **21**: p. 193-198.
153. Neal, E.G., et al., *The ketogenic diet for the treatment of childhood epilepsy: a randomised controlled trial*. *Lancet Neurol*, 2008. **7**: p. 500-6.
154. Payne, N.E., et al., *The ketogenic and related diets in adolescents and adults - a review*. *Epilepsia*, 2011. **52**: p. 1941-8.
155. Groesbeck, D., R. Bluml, and E.H. Kossoff, *Long-term use of the ketogenic diet in the treatment of epilepsy*. *Dev Med Child Neurol*, 2006. **48**: p. 978-981.
156. Vining, E.P.G., et al., *Growth of children on the ketogenic diet*. *Dev Med Child Neurol*, 2002. **44**: p. 796-802.
157. Williams, S., et al., *Growth retardation in children with epilepsy on the ketogenic diet: a retrospective chart review*. *J Am Diet Assoc*. 2002. **102**: p. 405-407.
158. Liu, Y.-M.C., et al., *A prospective study: growth and nutritional status of children treated with the ketogenic diet*. *J Am Diet Assoc*, 2003. **103**: p. 707-12.
159. Peterson, S.J., et al., *Changes in growth and seizure reduction in children on the ketogenic diet as a treatment for intractable epilepsy*. *J Am Diet Assoc*, 2005. **105**(5): p. 718-25.
160. Neal, E.G., et al., *Growth of children on classical and medium-chain triglyceride ketogenic diets*. *Pediatrics*, 2008. **122**(2): p. e334-40.
161. Couch, S. and F. Schwarzman, *Growth and nutritional outcomes of children treated with the ketogenic diet*. *J Am Diet Assoc*, 1999. **99**: p. 1573-1575.
162. Kwiterovich, P.O., Jr., et al., *Effect of a high-fat ketogenic diet on plasma levels of lipids, lipoproteins, and apolipoproteins in children*. *JAMA*, 2003. **290**(7): p. 912-20.
163. Bergqvist, A.G.C., *Long-term monitoring of the ketogenic diet: Do's and Don'ts*. *Epilepsy Res*, 2011. **100**: p.261-6.
164. Worden, L.T., et al., *Is there an ideal way to discontinue the ketogenic diet?* *Epilepsy Research*, 2011. **95**(3): p. 232-236.
165. Stafstrom, C.F. and Rho, J., *Epilepsy and the Ketogenic Diet*. 2004, Totowa, New Jersey: Humana Press Inc.
166. Bough, K.J. and J.M. Rho, *Anticonvulsant mechanisms of the ketogenic diet*. *Epilepsia*, 2007. **48**(1): p. 43-58.
167. Amari, A., Grace, N.C., and Fisher, W.W., *Achieving and maintaining compliance with the ketogenic diet*. *J Appl Behav Anal*, 1995. **28**: p. 341-342.
168. Westman, E.C., et al., *Low-carbohydrate nutrition and metabolism*. *Am J Clin Nutr*, 2007. **86**(2): p. 276-84.
169. Owen, O.E., et al., *Protein, fat, and carbohydrate requirements during starvation: anaplerosis and cataplerosis*. *Am J Clin Nutr*, 1998. **68**(1): p. 12-34.
170. Lieberman, M.A. and A. Marks, *Marks' Basic Medical Biochemistry: A Clinical Approach*. 2013, China: Lippincott Williams and Wilkins.
171. Manninen, A.H., *Metabolic effects of the very-low-carbohydrate diets: misunderstood "villains" of human metabolism*. *J Int Soc Sports Nutr*, 2004. **1**(2): p. 7-11.
172. Hasselbalch, S.G., et al., *Brain metabolism during short-term starvation in humans*. *J Cereb Blood Flow Metab*, 1994. **14**(1): p. 125-31.

173. Owen, O.E., et al., *Brain metabolism during fasting*. J Clin Invest, 1967. **46**(10): p. 1589-95.
174. Cahill, G.F., Jr., *Starvation in man*. N Engl J Med, 1970. **282**(12): p. 668-75.
175. Phinney, S.D., et al., *The human metabolic response to chronic ketosis without caloric restriction: preservation of submaximal exercise capability with reduced carbohydrate oxidation*. Metabolism, 1983. **32**(8): p. 769-76.
176. Grieb, P., et al., *Long-term consumption of a carbohydrate-restricted diet does not induce deleterious metabolic effects*. Nutr Res, 2008. **28**(12): p. 825-33.
177. Foster, G.D., et al., *A randomized trial of a low-carbohydrate diet for obesity*. N Engl J Med, 2003. **348**(21): p. 2082-90.
178. Stern, L., et al., *The effects of low-carbohydrate versus conventional weight loss diets in severely obese adults: one-year follow-up of a randomized trial*. Ann Intern Med, 2004. **140**(10): p. 778-85.
179. Nielsen, J.V. and Joensson, E.A., *Low-carbohydrate diet in type 2 diabetes. Stable improvement of bodyweight and glycemic control during 22 months follow-up*. Nutr Metab (Lond), 2006. **3**: p. 22.
180. Nielsen, J.V. and Joensson, E.A., *Low-carbohydrate diet in type 2 diabetes: stable improvement of bodyweight and glycemic control during 44 months follow-up*. Nutr Metab (Lond), 2008. **5**: p. 14.
181. Davis, N.J., et al., *Comparative study of the effects of a 1-year dietary intervention of a low-carbohydrate diet versus a low-fat diet on weight and glycemic control in type 2 diabetes*. Diabetes Care, 2009. **32**(7): p. 1147-52.
182. Nehlig, A., et al., *The ketogenic diet has no effect on the expression of spike-and-wave discharges and nutrient transporters in genetic absence epilepsy rats from Strasbourg*. J Neurochem, 2009. **109 Suppl 1**: p. 207-13.
183. Maalouf, M., Rho, J.M., and Mattson, M.P., *The neuroprotective properties of calorie restriction, the ketogenic diet, and ketone bodies*. Brain Res Rev, 2009. **59**(2): p. 293-315.
184. Yamada, K.A., *Calorie restriction and glucose regulation*. Epilepsia, 2008. **49 Suppl 8**: p. 94-6.
185. Greene, A.E., et al., *Caloric restriction inhibits seizure susceptibility in epileptic EL mice by reducing blood glucose*. Epilepsia, 2001. **42**(11): p. 1371-1378.
186. Eagles, D.A., et al., *Calorie restriction of a high-carbohydrate diet elevates the threshold of PTZ-induced seizures to values equal to those seen with a ketogenic diet*. Epilepsy Res, 2003. **54**(1): p. 41-52.
187. Bough, K.J., P.A. Schwartzkroin, and J.M. Rho, *Calorie restriction and ketogenic diet diminish neuronal excitability in rat dentate gyrus in vivo*. Epilepsia, 2003. **44**(6): p. 752-60.
188. Mantis, J.G., et al., *Management of multifactorial idiopathic epilepsy in EL mice with caloric restriction and the ketogenic diet: role of glucose and ketone bodies*. Nutr Metab (Lond), 2004. **1**(1): p. 11.
189. Freeman, J.M., J.B. Freeman, and M.T. Kelly, *The Ketogenic Diet: A Treatment for Epilepsy*, 2000, Demos: New York. p. 236.
190. Freeman, J. and Vining, E.P.G., *Seizures decrease rapidly after fasting: preliminary studies of the ketogenic diet*. Arch Pediatr Adolesc Med, 1999. **153**: p. 946-949.
191. Hartman, A.L., et al., *Seizure tests distinguish intermittent fasting from the ketogenic diet*. Epilepsia, 2010. **51**(8): p. 1395-1402.
192. Kennedy, A.R., et al., *A high-fat, ketogenic diet induces a unique metabolic state in mice*. Am J Physiol Endocrinol Metab, 2007. **292**(6): p. E1724-39.

193. Gilbert, D.L., Pyzik, P.L., and Freeman, J.M., *The ketogenic diet: seizure control correlates better with serum beta-hydroxybutyrate than with urine ketones*. J Child Neurol, 2000. **15**: p. 787-790.
194. Thavendiranathan, P., et al., *The MCT ketogenic diet: effects on animal seizure models*. Exp Neurol, 2000. **161**(2): p. 696-703.
195. Keith, H.M., *Factors Influencing Experimentally Produced Convulsions*. Arch of Neur and Psychiat, 1933. **29**(1): p. 148-154.
196. Rho, J.M., et al., *Acetoacetate, acetone, and dibenzylamine (a contaminant in L-(+)-beta-hydroxybutyrate) exhibit direct anticonvulsant actions in vivo*. Epilepsia, 2002. **43**(4): p. 358-61.
197. Likhodii, S.S., et al., *Anticonvulsant properties of acetone, a brain ketone elevated by the ketogenic diet*. Ann Neurol, 2003. **54**(2): p. 219-26.
198. Gasior, M., et al., *The anticonvulsant activity of acetone, the major ketone body in the ketogenic diet, is not dependent on its metabolites acetol, 1,2-propanediol, methylglyoxal, or pyruvic acid*. Epilepsia, 2007. **48**(4): p. 793-800.
199. Ma, W., Berg, J., and Yellen, G., *Ketogenic diet metabolites reduce firing in central neurons by opening K(ATP) channels*. J Neurosci, 2007. **27**(14): p. 3618-25.
200. Juge, N., et al., *Metabolic control of vesicular glutamate transport and release*. Neuron, 2010. **68**(1): p. 99-112.
201. Maalouf, M., et al., *Ketones inhibit mitochondrial production of reactive oxygen species production following glutamate excitotoxicity by increasing NADH oxidation*. Neuroscience, 2007. **145**(1): p. 256-64.
202. Noh, H.S., et al., *Acetoacetate protects neuronal cells from oxidative glutamate toxicity*. J Neurosci Res, 2006. **83**(4): p. 702-9.
203. Tanner, G.R., et al., *Single K ATP channel opening in response to action potential firing in mouse dentate granule neurons*. J Neurosci, 2011. **31**(23): p. 8689-96.
204. Seymour, K.J., et al., *Identification of cerebral acetone by 1H-MRS in patients with epilepsy controlled by ketogenic diet*. MAGMA, 1999. **8**(1): p. 33-42.
205. Thio, L.L., Wong, M., and Yamada, K.A., *Ketone bodies do not directly alter excitatory or inhibitory hippocampal synaptic transmission*. Neurology, 2000. **54**(2): p. 325-31.
206. Donevan, S.D., et al., *Voltage-dependent block of N-methyl-D-aspartate receptors by the novel anticonvulsant dibenzylamine, a bioactive constituent of L-(+)-beta-hydroxybutyrate*. Epilepsia, 2003. **44**(10): p. 1274-9.
207. Samoilova, M., et al., *Chronic in vitro ketosis is neuroprotective but not anti-convulsant*. J Neurochem, 2010. **113**(4): p. 826-35.
208. Than, K.D., et al., *Can you predict an immediate, complete, and sustained response to the ketogenic diet?* Epilepsia, 2005. **46**: p. 580-2.
209. Groesbeck, D., Bluml, R., and Kossoff, E.H., *Long-term use of the ketogenic diet in the treatment of epilepsy*. Dev Med Child Neurol, 2006. **48**: p. 978-981.
210. Pfeifer, H.H. and Thiele, E.A., *Low-glycemic-index treatment: A liberalized ketogenic diet for treatment of intractable epilepsy*. Neurology, 2005. **65**: p. 1810-1812.
211. Dressler, A., et al., *Long-term outcome and tolerability of the ketogenic diet in drug-resistant childhood epilepsy - The Austrian experience*. Seizure, 2010. **19**: p. 404-408.
212. Fraser, D.D., et al., *Elevated polyunsaturated fatty acids in blood serum obtained from children on the ketogenic diet*. Neurology, 2003. **60**: p. 1026-1029.
213. Hassan, A.M., et al., *Ketogenic diet in the treatment of refractory epilepsy in childhood*. Pediatr Neurol, 1999. **21**: p. 548-552.

214. Bough, K.J., Chen, R.S., and Eagles, D.A., *Path analysis shows that increasing ketogenic ratio, but not beta-hydroxybutyrate, elevates seizure threshold in the rat.* Dev Neurosci, 1999. **21**(3-5): p. 400-406.
215. Likhodii, S.S., et al., *Dietary fat, ketosis, and seizure resistance in rats on the ketogenic diet.* Epilepsia, 2000. **41**(11): p. 1400-10.
216. Hartman, A.L., et al., *Efficacy of the ketogenic diet in the 6-Hz seizure test.* Epilepsia, 2008. **49**(2): p. 334-339.
217. Nylen, K., et al., *A comparison of the ability of a 4:1 ketogenic diet and a 6.3:1 ketogenic diet to elevate seizure thresholds in adult and young rats.* Epilepsia, 2005. **46**(8): p. 1198-204.
218. Harney, J.P., et al., *Effects of acute inhibition of fatty acid oxidation on latency to seizure and concentrations of betahydroxybutyrate in plasma of rats maintained on calorie restriction and/or the ketogenic diet.* Epilepsy Res, 2002. **49**(3): p. 239-246.
219. Todorova, M.T., et al., *The ketogenic diet inhibits epileptogenesis in EL mice: a genetic model for idiopathic epilepsy.* Epilepsia, 2000. **41**(8): p. 933-940.
220. Stafstrom, C.E., Roopra, A., and Sutula, T.P., *Seizure suppression via glycolysis inhibition with 2-deoxy-D-glucose (2DG).* Epilepsia, 2008. **49 Suppl 8**: p. 97-100.
221. Greene, A.E., Todorova, M.T., and Seyfried, T.N., *Perspectives on the metabolic management of epilepsy through dietary reduction of glucose and elevation of ketone bodies.* J Neurochem, 2003. **86**(3): p. 529-37.
222. Panayiotopoulos, C.P., *Atlas of Epilepsies.* Vol. 3. 2010, London: Springer.
223. Li, X., Yokono, K., and Okada, Y., *Phosphofructokinase, a glycolytic regulatory enzyme has a crucial role for maintenance of synaptic activity in guinea pig hippocampal slices.* Neurosci Lett, 2000. **294**(2): p. 81-4.
224. Huttenlocher, P.R., *Ketonemia and seizures: metabolic and anticonvulsant effects of two ketogenic diets in childhood epilepsy.* Pediatr Res, 1976. **10**(5): p. 536-40.
225. Danial, N.N., *BAD: undertaker by night, candyman by day.* Oncogene, 2008. **27 Suppl 1**: p. S53-70.
226. Giménez-Cassina, A., et al., *BAD-dependent regulation of fuel metabolism and KATP channel activity confers resistance to epileptic seizures.* Neuron, 2012. **74**(4): p. 719-730.
227. Samala, R., Willis, S., and Borges, K., *Anticonvulsant profile of a balanced ketogenic diet in acute mouse seizure models.* Epilepsy Res, 2008. **81**(2): p. 119-127.
228. Mantis, J.G., et al., *Management of multifactorial idiopathic epilepsy in EL mice with caloric restriction and the ketogenic diet: role of glucose and ketone bodies.* Nutr Metab (Lond), 2004. **19**(1): p. 11.
229. Coppola, G., et al., *Low glycemic index diet in children and young adults with refractory epilepsy: First Italian experience.* Seizure, 2011. **20**: p. 526-528.
230. Bergqvist, A., et al., *Predictive power of first morning glucose and the ketogenic diet* Neuropediatrics, 2007. **38**(4): p. 193-196.
231. Nowotny, B., et al., *Precision and accuracy of blood glucose measurements using three different instruments.* Diabet Med, 2012. **29**(2): p. 260-5.
232. Kawamura, M., Jr., Ruskin, D.N., and Masino, S.A., *Metabolic autocrine regulation of neurons involves cooperation among pannexin hemichannels, adenosine receptors, and KATP channels.* J Neurosci, 2010. **30**(11): p. 3886-95.
233. Masino, S. and J. Rho, *Mechanisms of Ketogenic Diet Action.* In: Noebels JL, Avoli M, Rogawski MA, et al., editors. *Jasper's Basic Mechanisms of the Epilepsies [Internet]. 4th edition.*, 2012, Bethesda (MD): National Center for Biotechnology Information (US) .
234. Gasior, M., et al., *Anticonvulsant and proconvulsant actions of 2-deoxy-D-glucose.* Epilepsia, 2010. **51**(8): p. 1385-94.

235. Stafstrom, C.E., et al., *Anticonvulsant and antiepileptic actions of 2-deoxy-D-glucose in epilepsy models*. Ann Neurol, 2009. **65**(4): p. 435-47.
236. Lian, X.Y., F.A. Khan, and J.L. Stringer, *Fructose-1,6-bisphosphate has anticonvulsant activity in models of acute seizures in adult rats*. J Neurosci, 2007. **27**(44): p. 12007-11.
237. Garriga-Canut, M., et al., *2-Deoxy-D-glucose reduces epilepsy progression by NRSF-CtBP-dependent metabolic regulation of chromatin structure*. Nat Neurosci, 2006. **9**: p. 1382-7.
238. Yao, J., et al., *2-Deoxy-D-glucose treatment induces ketogenesis, sustains mitochondrial function, and reduces pathology in female mouse model of Alzheimer's disease*. PLoS One, 2011. **6**(7): p. e21788.
239. Voskuyl, R.A., et al., *Anticonvulsant effect of polyunsaturated fatty acids in rats, using the cortical stimulation model*. Eur J Pharmacol, 1998. **341**(2-3): p. 145-52.
240. Lauritzen, I., et al., *Polyunsaturated fatty acids are potent neuroprotectors*. EMBO J, 2000. **19**(8): p. 1784-93.
241. Yehuda, S., Carasso, R.L., and Mostofsky, D.I., *Essential fatty acid preparation (SR-3) raises the seizure threshold in rats*. Eur J Pharmacol, 1994. **254**(1-2): p. 193-8.
242. Taha, A.Y., et al., *Seizure resistance in fat-1 transgenic mice endogenously synthesizing high levels of omega-3 polyunsaturated fatty acids*. J Neurochem, 2008. **105**(2): p. 380-8.
243. Porta, N., et al., *Anticonvulsant effects of linolenic acid are unrelated to brain phospholipid cell membrane compositions*. Epilepsia, 2009. **50**(1): p. 65-71.
244. Blondeau, N., et al., *Polyunsaturated fatty acids induce ischemic and epileptic tolerance*. Neuroscience, 2002. **109**(2): p. 231-41.
245. Rabinovitz, S., D.I. Mostofsky, and S. Yehuda, *Anticonvulsant efficiency, behavioral performance and cortisol levels: a comparison of carbamazepine (CBZ) and a fatty acid compound (SR-3)*. Psychoneuroendocrinology, 2004. **29**(2): p. 113-24.
246. Taha, A.Y., et al., *Dose-dependent anticonvulsant effects of linoleic and alpha-linolenic polyunsaturated fatty acids on pentylenetetrazol induced seizures in rats*. Epilepsia, 2009. **50**(1): p. 72-82.
247. Taha, A.Y., et al., *Acute administration of docosahexaenoic acid increases resistance to pentylenetetrazol-induced seizures in rats*. Epilepsy and Behavior, 2010. **17**(3): p. 336-43.
248. Willis, S., et al., *Eicosapentaenoic and docosahexaenoic acids are not anticonvulsant or neuroprotective in acute mouse seizure models*. Epilepsia, 2009. **50**(1): p. 138-42.
249. Taha, A.Y., et al., *Lack of benefit of linoleic and alpha-linolenic polyunsaturated fatty acids on seizure latency, duration, severity or incidence in rats*. Epilepsy Res, 2006. **71**(1): p. 40-6.
250. Schlanger, S., Shinitzky, M., and Yam, D., *Diet enriched with omega-3 fatty acids alleviates convulsion symptoms in epilepsy patients*. Epilepsia, 2002. **43**(1): p. 103-4.
251. Yuen, A.W., et al., *Omega-3 fatty acid supplementation in patients with chronic epilepsy: a randomized trial*. Epilepsy Behav, 2005. **7**(2): p. 253-8.
252. Bromfield, E., et al., *A randomized trial of polyunsaturated fatty acids for refractory epilepsy*. Epilepsy Behav, 2008. **12**(1): p. 187-90.
253. DeGiorgio, C.M., et al., *n-3 fatty acids (fish oil) for epilepsy, cardiac risk factors, and risk of SUDEP: clues from a pilot, double-blind, exploratory study*. Epilepsy Behav, 2008. **13**(4): p. 681-4.

254. Varga, T., Czimmerer, Z., and Nagy, L., *PPARs are a unique set of fatty acid regulated transcription factors controlling both lipid metabolism and inflammation*. Biochim Biophys Acta, 2011. **1812**(8): p. 1007-22.
255. Lin, Q., et al., *Ligand selectivity of the peroxisome proliferator-activated receptor alpha*. Biochemistry, 1999. **38**(1): p. 185-90.
256. Cullingford, T.E., *The ketogenic diet; fatty acids, fatty acid-activated receptors and neurological disorders*. Prostaglandins Leukot Essent Fatty Acids, 2004. **70**(3): p. 253-64.
257. Cullingford, T.E., Eagles, D.A., and Sato, H., *The ketogenic diet upregulates expression of the gene encoding the key ketogenic enzyme mitochondrial 3-hydroxy-3-methylglutaryl-CoA synthase in rat brain*. Epilepsy Res, 2002. **49**: p. 99-107.
258. Kersten, S., et al., *Peroxisome proliferator-activated receptor alpha mediates the adaptive response to fasting*. J Clin Invest, 1999. **103**(11): p. 1489-98.
259. Xu, X., et al., *Attenuation of phospholipid signaling provides a novel mechanism for the action of valproic acid*. Eukaryotic Cell, 2007. **6**(6): p. 899-906.
260. Villarroya, F., Iglesias, R., and Giral, M., *PPARs in the Control of Uncoupling Proteins Gene Expression*. PPAR Res, 2007. **2007**: p. 74364.
261. Wu, Z., et al., *Mechanisms controlling mitochondrial biogenesis and respiration through the thermogenic coactivator PGC-1*. Cell, 1999. **98**(1): p. 115-24.
262. Ledesma, A., de Lacoba, M.G., and Rial, E., *The mitochondrial uncoupling proteins*. Genome Biology, 2002. **3**(12): p. REVIEWS3015.
263. Veech, R.L., *The therapeutic implications of ketone bodies: the effects of ketone bodies in pathological conditions: ketosis, ketogenic diet, redox states, insulin resistance, and mitochondrial metabolism*. Prostaglandins Leukot Essent Fatty Acids, 2004. **70**(3): p. 309-19.
264. Diano, S., et al., *Uncoupling protein 2 prevents neuronal death including that occurring during seizures: a mechanism for preconditioning*. Endocrinology, 2003. **144**(11): p. 5014-21.
265. Conti, B., et al., *Uncoupling protein 2 protects dopaminergic neurons from acute 1,2,3,6-methyl-phenyl-tetrahydropyridine toxicity*. J Neurochem, 2005. **93**(2): p. 493-501.
266. Andrews, Z.B., et al., *Uncoupling protein-2 is critical for nigral dopamine cell survival in a mouse model of Parkinson's disease*. J Neurosci, 2005. **25**(1): p. 184-91.
267. Sullivan, P.G., et al., *Mitochondrial uncoupling protein-2 protects the immature brain from excitotoxic neuronal death*. Ann Neurol, 2003. **53**(6): p. 711-7.
268. Sullivan, P.G., et al., *The ketogenic diet increases mitochondrial uncoupling protein levels and activity*. Ann Neurol, 2004. **55**(4): p. 576-80.
269. Andrews, Z.B., Diano, S., and Horvath, T.L., *Mitochondrial uncoupling proteins in the CNS: in support of function and survival*. Nat Rev Neurosci, 2005. **6**(11): p. 829-40.
270. Shimazu, T., et al., *Suppression of oxidative stress by beta-hydroxybutyrate, an endogenous histone deacetylase inhibitor*. Science, 2013. **339**(6116): p. 211-4.
271. Taha, A.Y., Burnham, W.M., and Auvin, S., *Polyunsaturated fatty acids and epilepsy*. Epilepsia, 2010. **51**(8): p. 1348-58.
272. Dekaban, A.S., *Plasma lipids in epileptic children treated with the high fat diet*. Arch Neurol, 1966. **15**(2): p. 177-84.
273. Chang, P., et al., *Seizure control by ketogenic diet-associated medium chain fatty acids*. Neuropharmacology, 2012.
274. Hughes, S.D., et al., *The ketogenic diet component decanoic acid increases mitochondrial citrate synthase and complex I activity in neuronal cells*. J Neurochem, 2014.

275. Haidukewych, D., Forsythe, W.I., and Sills, M., *Monitoring octanoic and decanoic acids in plasma from children with intractable epilepsy treated with medium-chain triglyceride diet*. Clin Chem, 1982. **28**(4 Pt 1): p. 642-5.
276. Sills, M.A., Forsythe, W.I., and Haidukewych, D., *Role of octanoic and decanoic acids in the control of seizures*. Arch Dis Child, 1986. **61**(12): p. 1173-7.
277. Dell, C.A., et al., *Lipid and fatty acid profiles in rats consuming different high-fat ketogenic diets*. Lipids, 2001. **36**(4): p. 373-8.
278. Ramirez-Camacho, A., et al., *Experiencia con la dieta cetogénica como tratamiento en la epilepsia refractaria. [Experience with ketogenic diet as treatment for refractory epilepsy]*. Rev Neurologia, 2011. **53**: p. 524-530.
279. Schwartz, R.H., et al., *Ketogenic diets in the treatment of epilepsy: short-term clinical effects*. Dev Med Child Neurol, 1989. **31**: p. 145-151.
280. Kim do, Y., et al., *Ketone bodies are protective against oxidative stress in neocortical neurons*. J Neurochem, 2007. **101**(5): p. 1316-26.
281. Kim do, Y., Vallejo, J., and Rho, J.M., *Ketones prevent synaptic dysfunction induced by mitochondrial respiratory complex inhibitors*. J Neurochem, 2010. **114**(1): p. 130-41.
282. Jarrett, S.G., et al., *The ketogenic diet increases mitochondrial glutathione levels*. J Neurochem, 2008. **106**(3): p. 1044-51.
283. Stafford, P., et al., *The ketogenic diet reverses gene expression patterns and reduces reactive oxygen species levels when used as an adjuvant therapy for glioma*. Nutr Metab, 2010. **7**: p. 74.
284. Mori, A., et al., *Natural antioxidants may prevent posttraumatic epilepsy: a proposal based on experimental animal studies*. Acta Medica Okayama, 2004. **58**(3): p. 111-8.
285. Cardenas-Rodriguez, N., et al., *Role of oxidative stress in refractory epilepsy: evidence in patients and experimental models*. Int J Mol Sci, 2013. **14**(1): p. 1455-76.
286. Dringen, R., *Metabolism and functions of glutathione in brain*. Prog Neurobiol, 2000. **62**(6): p. 649-71.
287. Ziegler, D.R., et al., *Ketogenic diet increases glutathione peroxidase activity in rat hippocampus*. Neurochem Res, 2003. **28**(12): p. 1793-7.
288. Lutas, A. and Yellen, G., *The ketogenic diet: metabolic influences on brain excitability and epilepsy*. Trends Neurosci, 2013. **36**(1): p. 32-40.
289. DeVivo, D.C., et al., *Chronic ketosis and cerebral metabolism*. Ann Neurol, 1978. **3**(4): p. 331-37.
290. Noh, H.S., et al., *A cDNA microarray analysis of gene expression profiles in rat hippocampus following a ketogenic diet*. Brain Res Mol Brain Res, 2004. **129**: p. 80-7.
291. Bough, K.J., et al., *Mitochondrial biogenesis in the anticonvulsant mechanism of the ketogenic diet*. Ann Neurol, 2006. **60**(2): p. 223-35.
292. Nylen, K., et al., *The effects of a ketogenic diet on ATP concentrations and the number of hippocampal mitochondria in Aldh5a1(-/-) mice*. Biochim Biophys Acta, 2009. **1790**(3): p. 208-12.
293. Ahola-Erkkila, S., et al., *Ketogenic diet slows down mitochondrial myopathy progression in mice*. Hum Mol Genet, 2010. **19**(10): p. 1974-84.
294. Pan, J.W., et al., *Ketosis and epilepsy: 31P spectroscopic imaging at 4.1 T*. Epilepsia, 1999. **40**(6): p. 703-7.
295. Yellen, G., *Ketone bodies, glycolysis, and KATP channels in the mechanism of the ketogenic diet*. Epilepsia, 2008. **49 Suppl 8**: p. 80-2.

296. McNally, M.A. and Hartman, A.L., *Ketone bodies in epilepsy*. Journal of neurochemistry, 2012. **121**(1): p. 28-35.
297. Li, D.P., Chen, S.R., and Pan, H.L., *Adenosine inhibits paraventricular pre-sympathetic neurons through ATP-dependent potassium channels*. J Neurochem, 2010. **113**(2): p. 530-42.
298. Kirsch, G.E., et al., *Coupling of ATP-sensitive K⁺ channels to A1 receptors by G proteins in rat ventricular myocytes*. Am J Physiol, 1990. **259**(3 Pt 2): p. H820-6.
299. Dart, C. and Standen, N.B., *Adenosine-activated potassium current in smooth muscle cells isolated from the pig coronary artery*. J Physiol, 1993. **471**: p. 767-86.
300. Li, Y., et al., *Adenosine modulates the excitability of layer II stellate neurons in entorhinal cortex through A1 receptors*. Hippocampus, 2011. **21**(3): p. 265-80.
301. Gouder, N., Fritschy, J.M., and Boison, D., *Seizure suppression by adenosine A1 receptor activation in a mouse model of pharmacoresistant epilepsy*. Epilepsia, 2003. **44**(7): p. 877-85.
302. Masino, S.A., et al., *A ketogenic diet suppresses seizures in mice through adenosine A(1) receptors*. J Clin Invest, 2011. **121**(7): p. 2679-83.
303. Yudkoff, M., et al., *Ketosis and brain handling of glutamate, glutamine, and GABA*. Epilepsia, 2008. **49 Suppl 8**: p. 73-5.
304. Hartman, A.L., et al., *The neuropharmacology of the ketogenic diet*. Pediatr Neurol, 2007. **36**(5): p. 281-92.
305. Cheng, C.M., et al., *Caloric restriction augments brain glutamic acid decarboxylase-65 and -67 expression*. J Neurosci Res, 2004. **77**(2): p. 270-6.
306. Suzuki, Y., et al., *Beta-hydroxybutyrate alters GABA-transaminase activity in cultured astrocytes*. Brain Res, 2009. **1268**: p. 17-23.
307. Yudkoff, M., et al., *Brain amino acid metabolism and ketosis*. Journal of Neuroscience Research, 2001. **66**(2): p. 272-81.
308. Al-Mudallal, A.S., et al., *Diet-induced ketosis does not cause cerebral acidosis*. Epilepsia, 1996. **37**(3): p. 258-61.
309. Melo, T.M., Nehlig, A., and Sonnewald, U., *Neuronal-glial interactions in rats fed a ketogenic diet*. Neurochem Int, 2006. **48**(6-7): p. 498-507.
310. Yudkoff, M., et al., *Response of brain amino acid metabolism to ketosis*. Neurochem Int, 2005. **47**(1-2): p. 119-28.
311. Cantello, R., et al., *Ketogenic Diet: Electrophysiological Effects on the Normal Human Cortex*. Epilepsia, 2007. **48**(9): p. 1756-1763.
312. Wang, Z.J., et al., *In vivo measurement of brain metabolites using two-dimensional double-quantum MR spectroscopy—exploration of GABA levels in a ketogenic diet*. Magn Reson Med, 2003. **49**(4): p. 615-619.
313. Dahlin, M., et al., *The ketogenic diet influences the levels of excitatory and inhibitory amino acids in the CSF in children with refractory epilepsy*. Epilepsy Res, 2005. **64**(3): p. 115-125.
314. Weinshenker, D. and Szot, P., *The role of catecholamines in seizure susceptibility: new results using genetically engineered mice*. Pharmacol Ther, 2002. **94**(3): p. 213-33.
315. Weinshenker, D., *The contribution of norepinephrine and orexigenic neuropeptides to the anticonvulsant effect of the ketogenic diet*. Epilepsia, 2008. **49 Suppl 8**: p. 104-7.
316. Szot, P., et al., *Norepinephrine is required for the anticonvulsant effect of the ketogenic diet*. Brain Res Dev Brain Res, 2001. **129**(2): p. 211-4.
317. Weinshenker, D., et al., *Genetic comparison of seizure control by norepinephrine and neuropeptide Y*. J Neurosci, 2001. **21**(19): p. 7764-9.

318. El Bahh, B., et al., *The anti-epileptic actions of neuropeptide Y in the hippocampus are mediated by Y and not Y receptors*. Eur J Neurosci, 2005. **22**(6): p. 1417-30.
319. Mazarati, A., Langel, U., and Bartfai, T., *Galanin: an endogenous anticonvulsant?* Neuroscientist, 2001. **7**(6): p. 506-17.
320. Tabb, K., et al., *The ketogenic diet does not alter brain expression of orexigenic neuropeptides*. Epilepsy Res, 2004. **62**(1): p. 35-9.
321. Harvey, J., *Leptin regulation of neuronal excitability and cognitive function*. Curr Opin Pharmacol, 2007. **7**(6): p. 643-7.
322. Obeid, M., et al., *Neuroprotective effects of leptin following kainic acid-induced status epilepticus*. Epilepsy Behav, 2010. **19**(3): p. 278-83.
323. Park, S., et al., *A ketogenic diet impairs energy and glucose homeostasis by the attenuation of hypothalamic leptin signaling and hepatic insulin signaling in a rat model of non-obese type 2 diabetes*. Exp Biol Med (Maywood), 2011. **236**(2): p. 194-204.
324. Thio, L.L., et al., *Leptin contributes to slower weight gain in juvenile rodents on a ketogenic diet*. Pediatr Res, 2006. **60**(4): p. 413-7.
325. Rho, J.M. and Stafstrom, C.E., *The ketogenic diet: what has science taught us?* Epilepsy Res, 2012. **100**(3): p. 210-7.
326. Willis, S., et al., *Anticonvulsant effects of a triheptanoin diet in two mouse chronic seizure models*. Neurobiol Dis, 2010. **40**(3): p. 565-72.
327. Aso, E., et al., *Triheptanoin supplementation to ketogenic diet curbs cognitive impairment in APP/PS1 mice used as a model of familial Alzheimer's disease*. Curr Alzheimer Res, 2012.
328. Lieberman, M. A. and Marks, A., *Synthesis and degradation of amino acids*, in *Mark's Basic Medical Biochemistry*, Lieberman, M.A. and Marks, A., Editors. 2012, Lippincott Williams & Wilkins: China. p. 731-751.
329. Evangelidou, A., Spilioti, M., and Doulioglou, V., *Branched chain amino acids as adjunctive therapy to ketogenic diet in epilepsy: pilot study and hypothesis*. J Child Neurol, 2009. **24**: p. 1268-1272.
330. Wang, D., et al., *Glut-1 deficiency syndrome: clinical, genetic, and therapeutic aspects*. Ann Neurol, 2005. **57**(1): p. 111-8.
331. Weber, T.A., Antognetti, M.R., and Stacpoole, P.W., *Caveats when considering ketogenic diets for the treatment of pyruvate dehydrogenase complex deficiency*. J Pediatr, 2000. **138**(3): p. 390-395.
332. Di Pisa, V., et al., *Case report of pyruvate dehydrogenase deficiency with unusual increase of fats during ketogenic diet treatment*. J Child Neurol, 2012. **27**(12): p. 1593-6.
333. Baranano, K.W. and A.L. Hartman, *The ketogenic diet: uses in epilepsy and other neurologic illnesses*. Curr Treat Options Neurol, 2008. **10**(6): p. 410-9.
334. Thammongkol, S., et al., *Efficacy of the ketogenic diet: Which epilepsies respond?* Epilepsia, 2012. **53**: p. e55-59.
335. Kossoff, E., *The fat is in the fire: ketogenic diet for refractory status epilepticus*. Epilepsy Curr, 2011. **11**(3): p. 88-9.
336. Nabbout, R., et al., *Efficacy of ketogenic diet in severe refractory status epilepticus initiating fever induced refractory epileptic encephalopathy in school age children (FIRES)*. Epilepsia, 2010. **51**(10): p. 2033-7.
337. Mikaeloff, Y., et al., *Devastating epileptic encephalopathy in school-aged children (DESC): a pseudo encephalitis*. Epilepsy Res, 2006. **69**(1): p. 67-79.
338. Oguni, H., et al., *Treatment and long-term prognosis of myoclonic-astatic epilepsy of early childhood*. Neuropediatrics, 2002. **33**(3): p. 122-32.

339. Caraballo, R.H., et al., *Ketogenic diet in patients with myoclonic-astatic epilepsy*. *Epileptic Disord*, 2006. **8**: p. 151-5.
340. Kilaru, S. and A.G.C. Bergqvist, *Current treatment of myoclonic astatic epilepsy: clinical experience at the Children's Hospital of Philadelphia*. *Epilepsia*, 2007. **48**: p. 1703-7.
341. Kang, H., et al., *Comparison of short- versus long-term ketogenic diet for intractable infantile spasms*. *Epilepsia*, 2011. **52**: p. 781-787.
342. Numis, A.L., et al., *The relationship of ketosis and growth to the efficacy of the ketogenic diet in infantile spasms*. *Epilepsy Res*, 2011. **96**(1-2): p. 172-5.
343. Kang, H.C., et al., *Safe and effective use of the ketogenic diet in children with epilepsy and mitochondrial respiratory chain complex defects*. *Epilepsia*, 2007. **48**(1): p. 82-8.
344. Caraballo, R.H. and N. Fejerman, *Dravet syndrome: a study of 53 patients*. *Epilepsy Res*, 2006. **70**(Suppl 1): p. S231-8.
345. Nabhout, R., et al., *Ketogenic diet also benefits Dravet syndrome patients receiving stiripentol: a prospective pilot study*. *Epilepsia*, 2011. **52**: p. e54-7.
346. Korff, C., et al., *Dravet syndrome (severe myoclonic epilepsy in infancy): a retrospective study of 16 patients*. *J Child Neurol*, 2007. **22**(2): p. 185-94.
347. Caraballo, R.H., *Nonpharmacologic treatments of Dravet syndrome: focus on the ketogenic diet*. *Epilepsia*, 2011. **52** (Suppl 2): p. 79-82.
348. Mullen, S.A., et al., *Glucose transporter 1 deficiency as a treatable cause of myoclonic astatic epilepsy*. *Arch Neurol*, 2011. **68**(9): p. 1152-5.
349. Arsov, T., et al., *Early onset absence epilepsy: 1 in 10 cases is caused by GLUT1 deficiency*. *Epilepsia*, 2012. **53**(12): p. e204-7.
350. Suls, A., et al., *Early-onset absence epilepsy caused by mutations in the glucose transporter GLUT1*. *Ann Neurol*, 2009. **66**(3): p. 415-9.
351. Schoeler, N.E., et al., *Can we predict a favourable response to Ketogenic Diet Therapies for drug-resistant epilepsy?* *Epilepsy Res*, 2013.
352. Freeman, J.M., et al., *The efficacy of the ketogenic diet-1998: a prospective evaluation of intervention in 150 children*. *Pediatrics*, 1998. **102**: p. 1358-1363.
353. Katyal, N.G., et al., *The ketogenic diet in refractory epilepsy: the experience of Children's Hospital of Pittsburgh*. *Clin Pediatr (Phila)*, 2000. **39**: p. 153-159.
354. Nordli, J.D.R., et al., *Experience with the ketogenic diet in infants*. *Pediatrics*, 2001. **108**: p. 129-133.
355. Coppola, G., et al., *The ketogenic diet in children , adolescents and young adults with refractory epilepsy: an Italian multicentric experience*. *Epilepsy Res*, 2002. **48**: p. 221-227.
356. Dimario, F.J. and Holland, J., *The ketogenic diet: a review of the experience at Connecticut Children's Medical Center*. *Pediatr Neurol*, 2002. **26**: p. 288-292.
357. Sharma, S., et al., *Seizure control and biochemical profile on the ketogenic diet in young children with refractory epilepsy - Indian experience*. *Seizure*, 2009. **18**: p. 446-9.
358. Coppola, G., et al., *Ketogenic diet for the treatment of catastrophic epileptic encephalopathies in childhood*. *Eur J Paediatr Neurol*, 2010. **14**: p. 229-234.
359. Caraballo, R., et al., *Long-term follow-up of the ketogenic diet for refractory epilepsy: multicenter Argentinean experience in 216 pediatric patients*. *Seizure*, 2011. **20**: p. 640-5.
360. Groomes, L.B., et al., *Do patients with absence epilepsy respond to ketogenic diets?* *J Child Neurol*, 2011. **26**: p. 160-165.
361. Moreno-Villares, J.M., et al., *La vuelta a la dieta cetogénica. ¿Qué papel desempeña en el tratamiento de las convulsiones infantiles refractarias?* [The return to the

- ketogenic diet. What role does it play in the treatment of refractory seizures of infancy?]. Rev Neurologia, 2001. 32: p. 1115-1119.*
362. Yum, M.K., et al., *Effect of a ketogenic diet on EEG: analysis of sample entropy. Seizure, 2008. 17: p. 561-6.*
 363. Kang, H., et al., *Efficacy and safety of the ketogenic diet for intractable childhood epilepsy: Korean multicentric experience. Epilepsia, 2005. 46: p. 272-279.*
 364. Hopkins, I. and Lynch, B.C., *Use of the ketogenic diet in epilepsy in childhood. Proc Aust Assoc Neurol, 1970. 7: p. 25-30.*
 365. Swink, T., Vining, E., and Freeman, J., *The ketogenic diet: 1997. Adv Pediatr, 1997. 44: p. 297-329.*
 366. Wirrell, E.C., et al., *Is a fast necessary when initiating the ketogenic diet? J Child Neurol, 2002. 17: p. 179-182.*
 367. Ros Pérez, P., et al., *Valoración de la efectividad de la dieta cetogénica con triglicéridos de cadena media (TCM), en el tratamiento de la epilepsia refractaria en niños. A propósito de una casuística [Evaluation of the effectiveness of the ketogenic diet with medium-chain triglycerides, in the treatment of refractory epilepsy in children. Apropos of a series of cases]. An Esp Pediatr, 1989. 30(3): p. 155-158.*
 368. Morrison, P.F., et al., *The influence of concurrent anticonvulsants on the efficacy of the ketogenic diet. Epilepsia, 2009. 50: p. 1999-2001.*
 369. Peterson, S.J., et al., *Changes in growth and seizure reduction in children on the ketogenic diet as a treatment for intractable epilepsy. J Am Diet Assoc, 2005. 105: p. 718-25.*
 370. Raju, K.N., et al., *Efficacy of 4:1 (classic) versus 2.5:1 ketogenic ratio diets in refractory epilepsy in young children: a randomized open labeled study. Epilepsy Res, 2011. 96(1-2): p. 96-100.*
 371. El-Rashidy, O.F., et al., *Modified Atkins diet vs classic ketogenic formula in intractable epilepsy. Acta Neurol Scand, 2013.*
 372. Panico, L.R., et al., *Dieta cetogénica en la epilepsia refractaria infantil: respuesta electroclínica, complicaciones y efectos secundarios [The ketogenic diet in infantile refractory epilepsy: electroclinical response, complications and secondary effects]. Rev Neurologia, 2000. 31(3): p. 212-220.*
 373. Freeman, J.M., Kossoff, E.H., and Hartman, A.L., *The Ketogenic Diet: One Decade Later. Pediatrics, 2007. 119: p. 535-543.*
 374. Nathan, J.K., et al., *Ketogenic diet in Indian children with uncontrolled epilepsy. Indian Pediatr, 2009. 46: p. 669-73.*
 375. Schwartz, R., Boyes, S., and Aynsley-Green, A., *Metabolic effects of three ketogenic diets in the treatment of severe epilepsy. Dev Med Child Neurol, 1989. 31: p. 152-160.*
 376. Musa-Veloso, K., et al., *Breath acetone predicts plasma ketone bodies in children with epilepsy on a ketogenic diet. Nutrition, 2006. 22: p. 1-8.*
 377. Tagliabue, A., et al., *Effects of the ketogenic diet on nutritional status, resting energy expenditure, and substrate oxidation in patients with medically refractory epilepsy: A 6-month prospective observational study. Clin Nutr, 2012. 31: p. 246-9.*
 378. Dahlin, M., Mansson, J.E., and Amark, P., *CSF levels of dopamine and serotonin, but not norepinephrine, metabolites are influenced by the ketogenic diet in children with epilepsy. Epilepsy Res, 2012. 99(1-2): p. 132-8.*
 379. Janaki, S., et al., *A clinical electroencephalographic correlation of seizures on a ketogenic diet. Ind J Med Res, 1976. 64(7): p. 1057-63.*

380. François, L.L., et al., *Le régime cétoène à visée anti-épileptique: son utilisation chez 29 enfants épileptiques [Ketogenic regime as anti-epileptic treatment: its use in 29 epileptic children]*. Arch Pediatr, 2003. **10**: p. 300-306.
381. Nehlig, A. and Pereira de Vasconcelos, A., *Glucose and ketone body utilization by the brain of neonatal rats*. Prog Neurobiol, 1993. **40**(2): p. 163-221.
382. Bough, K.J., et al., *Seizure resistance is dependent upon age and calorie restriction in rats fed a ketogenic diet*. Epilepsy Res, 1999. **35**(1): p. 21-28.
383. Bough, K.J. and Eagles, D.A., *A ketogenic diet increases the resistance to pentylenetetrazole-induced seizures in the rat*. Epilepsia, 1999. **40**(2): p. 138-43.
384. Bough, K.J., Yao, S.G., and Eagles, D.A., *Higher ketogenic diet ratios confer protection from seizures without neurotoxicity*. Epilepsy Res, 1999. **38**(1): p. 15-25.
385. Bough, K.J., Matthews, P.J., and Eagles, D.A., *A ketogenic diet has different effects upon seizures induced by maximal electroshock and by pentylenetetrazole infusion*. Epilepsy Res, 2000. **38**(2-3): p. 105-14.
386. Zarnowska, I., et al., *Pharmacodynamic and pharmacokinetic interactions between common antiepileptic drugs and acetone, the chief anticonvulsant ketone body elevated in the ketogenic diet in mice*. Epilepsia, 2009. **50**(5): p. 1132-1140.
387. Hung, S.I., et al., *Genetic susceptibility to carbamazepine-induced cutaneous adverse drug reactions*. Pharmacogenet Genomics, 2006. **16**(4): p. 297-306.
388. Chen, P., et al., *Carbamazepine-induced toxic effects and HLA-B*1502 screening in Taiwan*. N Engl J Med, 2011. **364**(12): p. 1126-33.
389. Arnaout, R., et al., *Predicting the cost and pace of pharmacogenomic advances: an evidence-based study*. Clin Chem, 2013. **59**(4): p. 649-57.
390. Rattanaipapong, W., et al., *Economic Evaluation of HLA-B*1502 Genotyping in Carbamazepine Induced Stevens-Johnson Syndrome (SJS) and Toxic Epidermal Necrosis (TEN)*. Draft report., 2012, Health Intervention and Technology Assessment Program: Nonthaburi.
391. Ferriss, J.S., et al., *Multi-gene expression predictors of single drug responses to adjuvant chemotherapy in ovarian carcinoma: predicting platinum resistance*. PLoS One, 2012. **7**(2): p. e30550.
392. Mutch, D.M., et al., *A distinct adipose tissue gene expression response to caloric restriction predicts 6-mo weight maintenance in obese subjects*. Am J Clin Nutr, 2011. **94**(6): p. 1399-409.
393. Timmons, J.A., et al., *Using molecular classification to predict gains in maximal aerobic capacity following endurance exercise training in humans*. J Appl Physiol, 2010. **108**(6): p. 1487-96.
394. Bough, K.J., et al., *Mitochondrial biogenesis in the anticonvulsant mechanism of the ketogenic diet*. Ann Neurol, 2006. **60**(2): p. 223-35.
395. Williams, R.B., et al., *The influence of genetic variation on gene expression*. Genome Research, 2007. **17**(12): p. 1707-16.
396. Grundberg, E., T. Kwan, and T.M. Pastinen, *Analysis of the impact of genetic variation on human gene expression*. Methods in Molecular Biology, 2010. **628**: p. 321-39.
397. Fu, J., et al., *Unraveling the regulatory mechanisms underlying tissue-dependent genetic variation of gene expression*. PLoS Genetics, 2012. **8**(1): p. e1002431.
398. Nangia, S., et al., *Is the ketogenic diet effective in specific epilepsy syndromes?* Epilepsy Res, 2012. **100**(3): p. 252-257.
399. Dutton, S.B.B. and Escayg, A., *Genetic influences on ketogenic diet efficacy*. Epilepsia, 2008. **49 Suppl 8**: p. 67-69.

400. Fenoglio-Simeone, K.A., et al., *Ketogenic diet treatment abolishes seizure periodicity and improves diurnal rhythmicity in epileptic Kcna1-null mice*. *Epilepsia*, 2009. **50**(9): p. 2027-2034.
401. Dutton, S.B.B., et al., *Protective effect of the ketogenic diet in Scn1a mutant mice*. *Epilepsia*, 2011. **52**: p. 2050-6.
402. Hagopian, K., et al., *Shc proteins influence the activities of enzymes involved in fatty acid oxidation and ketogenesis*. *Metabolism* 2012. **61**(12): p. 1703-13.
403. Raddatz, K., et al., *Deletion of protein kinase C ϵ in mice has limited effects on liver metabolite levels but alters fasting ketogenesis and gluconeogenesis*. *Diabetologia*, 2012. **55**: p. 2789-2793.
404. Oishi, K., et al., *PPAR α deficiency augments a ketogenic diet-induced circadian PAI-1 expression possibly through PPAR γ activation in the liver*. *Biochem Biophys Res Commun*, 2010. **401**: p. 313-8.
405. Badman, M.K., et al., *Fibroblast growth factor 21-deficient mice demonstrate impaired adaptation to ketosis*. *Endocrinology*, 2009. **150**(11): p. 4931-4940.
406. Hotta, Y., et al., *Fibroblast growth factor 21 regulates lipolysis in white adipose tissue but is not required for ketogenesis and triglyceride clearance in liver*. *Endocrinology*, 2009. **150**(10): p. 4625-4633.
407. Henderson, S.T., et al., *Study of the ketogenic agent AC-1202 in mild to moderate Alzheimer's disease: a randomized, double-blind, placebo-controlled, multicenter trial*. *Nutr Metab*, 2009. **6**: p. 31.
408. Reger, M.a., et al., *Effects of beta-hydroxybutyrate on cognition in memory-impaired adults*. *Neurobiol Aging*, 2004. **25**: p. 311-4.
409. Plank, M., et al., *A meta-analysis of caloric restriction gene expression profiles to infer common signatures and regulatory mechanisms*. *Mol BioSys*, 2012. **8**(4): p. 1339-1349.
410. Kobow, K., et al., *Deep sequencing reveals increased DNA methylation in chronic rat epilepsy*. *Acta Neuropathol*, 2013. **126**: p. 741-756.
411. Cheng, C.M., *A Ketogenic Diet Increases Brain Insulin-Like Growth Factor Receptor and Glucose Transporter Gene Expression*. *Endocrinology*, 2003. **144**: p. 2676-2682.
412. Milder, J.B., Liang, L.P., and Patel, M., *Acute oxidative stress and systemic Nrf2 activation by the ketogenic diet*. *Neurobiol Dis*, 2010. **40**: p. 238-44.
413. Badman, M.K., et al., *Hepatic fibroblast growth factor 21 is regulated by PPAR α and is a key mediator of hepatic lipid metabolism in ketotic states*. *Cell Metab*, 2007. **5**(6): p. 426-37.
414. Jeong, E.A., et al., *Ketogenic diet-induced peroxisome proliferator-activated receptor-gamma activation decreases neuroinflammation in the mouse hippocampus after kainic acid-induced seizures*. *Exp Neurol*, 2011. **232**(2): p. 195-202.
415. Mohamed, H.E., et al., *Biochemical effect of a ketogenic diet on the brains of obese adult rats*. *J Clin Neurosci*, 2010. **17**(7): p. 899-904.
416. Badman, M.K., et al., *A very low carbohydrate ketogenic diet improves glucose tolerance in ob/ob mice independently of weight loss*. *Am J Physiol Endocrinol Metab*, 2009. **297**(5): p. E1197-204.
417. Kim do, Y., et al., *Inflammation-mediated memory dysfunction and effects of a ketogenic diet in a murine model of multiple sclerosis*. *PLoS One*, 2012. **7**(5): p. e35476.
418. McDaniel, S.S., et al., *The ketogenic diet inhibits the mammalian target of rapamycin (mTOR) pathway*. *Epilepsia*, 2011. **52**: p. e7-11.
419. Ragsdale, D.S., *How do mutant Nav1.1 sodium channels cause epilepsy?* *Brain Res Rev*, 2008. **58**(1): p. 149-59.

420. Lee, A.L., et al., *Potassium channel gene therapy can prevent neuron death resulting from necrotic and apoptotic insults*. J Neurochem, 2003. **86**(5): p. 1079-88.
421. Brandon, N.J., et al., *Receptor for activated C kinase-1 facilitates protein kinase C-dependent phosphorylation and functional modulation of GABA(A) receptors with the activation of G-protein-coupled receptors*. J Neurosci, 2002. **22**(15): p. 6353-61.
422. Tamura, Y., S. Simizu, and H. Osada, *The phosphorylation status and anti-apoptotic activity of Bcl-2 are regulated by ERK and protein phosphatase 2A on the mitochondria*. FEBS Letters, 2004. **569**(1-3): p. 249-55.
423. Woo, Y.C., et al., *Fibroblast growth factor 21 as an emerging metabolic regulator: clinical perspectives*. Clin Endocrinol (Oxf), 2013. **78**(4): p. 489-96.
424. Hsueh, H., W. Pan, and A.J. Kastin, *The fasting polypeptide FGF21 can enter brain from blood*. Peptides, 2007. **28**(12): p. 2382-6.
425. Sarraf, D.A., et al., *Fibroblast growth factor 21 action in the brain increases energy expenditure and insulin sensitivity in obese rats*. Diabetes, 2010. **59**(7): p. 1817-24.
426. Langert, K.A., Von Zee, C.L., and Stubbs Jr, E.B., *Tumor Necrosis Factor-alpha Enhances CCL2 and ICAM-1 Expression in Peripheral Nerve Microvascular Endoneurial Endothelial Cells*. ASN Neuro, 2013. **5**(1): p. e00104.
427. Im, J.Y., et al., *Cyclooxygenase-2-dependent neuronal death proceeds via superoxide anion generation*. Free Radic Biol Med, 2006. **41**(6): p. 960-72.
428. Toscano, C.D., et al., *Altered GABAergic neurotransmission is associated with increased kainate-induced seizure in prostaglandin-endoperoxide synthase-2 deficient mice*. Brain Res Bull, 2008. **75**(5): p. 598-609.
429. Bordone, L. and Guarente, L., *Calorie restriction, SIRT1 and metabolism: understanding longevity*. Nat Rev Mol Cell Biol, 2005. **6**(4): p. 298-305.
430. Van der Auwera, I., et al., *A ketogenic diet reduces amyloid beta 40 and 42 in a mouse model of Alzheimer's disease*. Nutr Metab (Lond), 2005. **2**: p. 28.
431. Garcia-Roves, P., et al., *Raising plasma fatty acid concentration induces increased biogenesis of mitochondria in skeletal muscle*. Proc Natl Acad Sci USA., 2007. **104**(25): p. 10709-13.
432. Sparks, L.M., et al., *A high-fat diet coordinately downregulates genes required for mitochondrial oxidative phosphorylation in skeletal muscle*. Diabetes, 2005. **54**(7): p. 1926-33.
433. Motsinger-Reif, A.A., et al., *Genome-wide association studies in pharmacogenomics: successes and lessons*. Pharmacogenet Genomics, 2010.
434. McCormack, M., et al., *HLA-A*3101 and Carbamazepine-Induced Hypersensitivity Reactions in Europeans*. N Engl J Med, 2011. **364**(12): p. 1134-1143.
435. Gilissen, C., et al., *Exome sequencing identifies WDR35 variants involved in Sensenbrenner syndrome*. Am J Hum Genet, 2010. **87**(3): p. 418-423.
436. Ng, S.B., et al., *Exome sequencing identifies the cause of a mendelian disorder*. Nat Genet, 2010. **42**(1): p. 30-35.
437. Choi, M., et al., *Genetic diagnosis by whole exome capture and massively parallel DNA sequencing*. Proc Natl Acad Sci USA, 2009. **106**(45): p. 19096-19101.
438. Thompson, K., Kulkarni, J., and Sergejew, A.A., *Reliability and validity of a new Medication Adherence Rating Scale (MARS) for the psychoses*. Schizophr Res, 2000. **42**(3): p. 241-7.
439. Mahler, C., et al., *Assessing reported adherence to pharmacological treatment recommendations. Translation and evaluation of the Medication Adherence Report Scale (MARS) in Germany*. J Eval Clin Pract, 2010. **16**(3): p. 574-9.
440. Li, H. and Durbin, R., *Fast and accurate short read alignment with Burrows-Wheeler transform*. Bioinformatics, 2009. **25**(14): p. 1754-60.

441. Koboldt, D.C., et al., *VarScan: variant detection in massively parallel sequencing of individual and pooled samples*. Bioinformatics, 2009. **25**(17): p. 2283-5.
442. Rainen, L., et al., *Stabilization of mRNA expression in whole blood samples*. Clin Chem, 2002. **48**(11): p. 1883-1890.
443. Hisham Yahaya, B., et al., *Optimising the quality and integrity of RNA samples from bronchial airway tissues*. AsPac J. Mol. Biol. Biotechnol, 2011. **19**(1): p. 19-27.
444. Nachtomy, O., Shavit, A., and Yakhini, Z., *Gene expression and the concept of the phenotype*. Stud Hist Philos Biol Biomed Sci, 2007. **38**(1): p. 238-54.
445. Robinson, P.N., *Deep phenotyping for precision medicine*. Hum Mutat, 2012. **33**(5): p. 777-80.
446. Schulze, T.G. and McMahon, F.J., *Defining the phenotype in human genetic studies: forward genetics and reverse phenotyping*. Hum Hered, 2004. **58**(3-4): p. 131-8.
447. Tracy, R.P., 'Deep phenotyping': characterizing populations in the era of genomics and systems biology. Curr Opin Lipidol, 2008. **19**(2): p. 151-7.
448. Pullar, T., Kumar, S., and Feely, M., *Compliance in clinical trials*. Ann Rheum Dis, 1989. **48**(10): p. 871-5.
449. Feise, R.J., *Do multiple outcome measures require p-value adjustment?* BMC Med Res Methodol, 2002. **2**: p. 8.
450. Kline, P., *The handbook of psychological testing*. 2nd ed. 1999, London: Routledge.
451. Cortina, J.M., *What is coefficient alpha? An examination of theory and applications*. J Appl Psychol, 1993. **78**(1): p. 98-104.
452. Tavakol, M. and Dennick, R., *Making sense of Cronbach's alpha*. Int J Med Educat, 2011. **2**: p. 53-55.
453. Sheng, Y. and Sheng, Z., *Is coefficient alpha robust to non-normal data?* Front Psychol, 2012. **3**: p. 34.
454. Neal, E., *Dietary Treatment of Epilepsy: Practical Implementation of Ketogenic Therapy*. 2012: Wiley-Blackwell.
455. Manikandan, S., *Measures of dispersion*. J Pharmacol Pharmacother, 2011. **2**(4): p. 315-6.
456. Cramer, J. and Mattson, R., *Monitoring compliance with antiepileptic drug therapy, in Patient compliance in medical practice and clinical trials*, J. Cramer and Spilker, B., Editors. 1991, Raven Press: New York. p. 123-137.
457. Haynes, R.B., et al., *Can simple clinical measurements detect patient noncompliance?* Hypertension, 1980. **2**(6): p. 757-64.
458. Stephenson, B.J., et al., *The rational clinical examination. Is this patient taking the treatment as prescribed?* JAMA, 1993. **269**(21): p. 2779-81.
459. Osterberg, L. and Blaschke, T., *Adherence to medication*. N Engl J Med, 2005. **353**(5): p. 487-97.
460. Goldstein, D.E., et al., *Tests of glycemia in diabetes*. Diabetes Care, 2004. **27**(7): p. 1761-73.
461. Musa-Veloso, K., Likhodii, S.S., and Cunnane, S.C., *Breath acetone is a reliable indicator of ketosis in adults consuming ketogenic meals*. Am J Clin Nutr, 2002. **76**(1): p. 65-70.
462. Hoppe, C., Poepel, A., and Elger, C.E., *Epilepsy: accuracy of patient seizure counts*. Arch Neurol, 2007. **64**(11): p. 1595-1599.
463. Cuetter, A.C., *On the topic of accuracy of patient seizure counts*. Arch Neurol, 2008. **65**(6): p. 849-850.
464. Bush, W.S. and Moore, J.H., *Chapter 11: Genome-wide association studies*. PLoS Comput Biol, 2012. **8**(12): p. e1002822.
465. Ackerson, B.K., et al., *Agreement between medical record and parent report for evaluation of childhood febrile seizures*. Vaccine, 2013. **31**(27): p. 2904-9.

466. Asaka, Y. and Takada, S., *Comparing sleep measures of infants derived from parental reports in sleep diaries and acceleration sensors*. *Acta Paediatr*, 2011. **100**(8): p. 1158-63.
467. Blum, D.E., et al., *Patient awareness of seizures*. *Neurology*, 1996. **47**(1): p. 260-4.
468. Poochikian-Sarkissian, S., et al., *Patient awareness of seizures as documented in the epilepsy monitoring unit*. *Can J Neurosci Nurs*, 2009. **31**(4): p. 22-3.
469. Heo, K., et al., *Patient awareness of complex partial seizures*. *Epilepsia*, 2006. **47**(11): p. 1931-5.
470. Fisher, R.S., et al., *Seizure diaries for clinical research and practice: limitations and future prospects*. *Epilepsy Behav*, 2012. **24**(3): p. 304-10.
471. Pujar, S., et al., *Statistical process control (SPC)--a simple objective method for monitoring seizure frequency and evaluating effectiveness of drug interventions in refractory childhood epilepsy*. *Epilepsy Res*, 2010. **91**(2-3): p. 205-13.
472. Westover, M.B., et al., *Revising the "Rule of Three" for inferring seizure freedom*. *Epilepsia*, 2012. **53**(2): p. 368-376.
473. McLachlan, R.S., et al., *Health-related quality of life and seizure control in temporal lobe epilepsy*. *Annals of Neurology*, 1997. **41**(4): p. 482-489.
474. Birbeck, G.L., et al., *Seizure reduction and quality of life improvements in people with epilepsy*. *Epilepsia*, 2002. **43**(5): p. 535-538.
475. Leidy, N.K., et al., *Seizure frequency and the health-related quality of life of adults with epilepsy*. *Neurology*, 1999. **53**(1): p. 162-6.
476. Sabaz, M., et al., *The impact of epilepsy surgery on quality of life in children*. *Neurology*, 2006. **66**(4): p. 557-61.
477. Schiller, Y. and Najjar, Y., *Quantifying the response to antiepileptic drugs: effect of past treatment history*. *Neurology*, 2008. **70**(1): p. 54-65.
478. Kwan, P. and Brodie, M.J., *Early identification of refractory epilepsy*. *N Engl J Med*, 2000. **342**(5): p. 314-9.
479. Kwan, P., Schachter, S.C., and Brodie, M.J., *Drug-resistant epilepsy*. *N Engl J Med*, 2011. **365**: p. 919-926.
480. Krueger, D.A., et al., *Everolimus treatment of refractory epilepsy in tuberous sclerosis complex*. *Ann Neurol*, 2013.
481. Somerville, E.R., *Some treatments cause seizure aggravation in idiopathic epilepsies (especially absence epilepsy)*. *Epilepsia*, 2009. **50 Suppl 8**: p. 31-6.
482. Kossoff, E.H., et al., *Efficacy of the ketogenic diet for infantile spasms*. *Pediatrics*, 2002. **109**(5): p. 780-3.
483. Faught, E., *Topiramate in the treatment of partial and generalized epilepsy*. *Neuropsychiatr Dis Treat*, 2007. **3**(6): p. 811-21.
484. Choi, J.N., et al., *Renal stone associated with the ketogenic diet in a 5-year old girl with intractable epilepsy*. *Yonsei Med J*, 2010. **51**(3): p. 457-9.
485. Kang, H.C., et al., *Early- and Late-onset Complications of the Ketogenic Diet for Intractable Epilepsy*. *Epilepsia*, 2004. **45**: p. 1116-1123.
486. Furth, S.L., et al., *Risk factors for urolithiasis in children on the ketogenic diet*. *Pediatr Nephrol*, 2000. **15**(1-2): p. 125-8.
487. Takeoka, M., et al., *Concomitant treatment with topiramate and ketogenic diet in pediatric epilepsy*. *Epilepsia*, 2002. **43**: p. 1072-5.
488. Sancho, J., et al., *Changes in seizure severity and quality of life in patients with refractory partial epilepsy*. *Epilepsy Behav*, 2010. **19**(3): p. 409-13.
489. Vickrey, B.G., et al., *Relationships between seizure severity and health-related quality of life in refractory localization-related epilepsy*. *Epilepsia*, 2000. **41**(6): p. 760-4.

490. Adebayo, P.B., et al. *Seizure severity and health-related quality of life of adult Nigerian patients with epilepsy*. Acta Neurol Scand, 2013. **129**(2): p. 102-8.
491. Harden, C.L., et al., *The effect of seizure severity on quality of life in epilepsy*. Epilepsy Behav, 2007. **11**(2): p. 208-11.
492. Vaisleib, I.I., J.R. Buchhalter, and M.L. Zupanc, *Ketogenic diet: outpatient initiation, without fluid, or caloric restrictions*. Pediatr Neurol, 2004. **31**: p. 198-202.
493. Neal, E.G., et al., *The ketogenic diet as a treatment for children with intractable epilepsy: Early drop out rate and efficacy*. Epilepsia, 2005. **46**: p. 230.
494. Lambrechts, D., et al., *The ketogenic diet as a treatment option in adults with chronic refractory epilepsy: Efficacy and tolerability in clinical practice*. Epilepsy Behav, 2012. **23**: p. 310-4.
495. Duncan, J.S. and Sander, J.W., *The Chalfont Seizure Severity Scale*. J Neurol Neurosurg Psychiatry, 1991. **54**(10): p. 873-6.
496. O'Donoghue, M.F., Duncan, J.S., and Sander, J.W., *The National Hospital Seizure Severity Scale: a further development of the Chalfont Seizure Severity Scale*. Epilepsia, 1996. **37**(6): p. 563-71.
497. Baker, G.A., et al., *The development of a seizure severity scale as an outcome measure in epilepsy*. Epilepsy Res, 1991. **8**(3): p. 245-51.
498. Cramer, J.A., et al., *A method of quantification for the evaluation of antiepileptic drug therapy*. Neurology, 1983. **33**(3 Suppl 1): p. 26-37.
499. Zhu, M. and Zhao, S., *Candidate gene identification approach: progress and challenges*. Int J Biol Sci, 2007. **3**(7): p. 420-7.
500. Minihihane, A.M., *The genetic contribution to disease risk and variability in response to diet: where is the hidden heritability?* Proc Nutr Soc, 2013. **72**(1): p. 40-7.
501. Minihihane, A.M., et al., *ApoE genotype, cardiovascular risk and responsiveness to dietary fat manipulation*. Proc Nutr Soc, 2007. **66**(2): p. 183-97.
502. Charland, S.L., et al., *Patient knowledge of pharmacogenetic information improves adherence to statin therapy: Results of the additional kif6 risk offers better adherence to statins (akrobats) trial*. Coll Cardiol, 2012. **59**(13s1): p. E1848.
503. Halushka, M.K., et al., *Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis*. Nat Genet, 1999. **22**(3): p. 239-47.
504. Xu, J.Y., Xu, G.B., and Chen, S.L., *A new method for SNP discovery*. Biotechniques, 2009. **46**(3): p. 201-8.
505. Zhao, H., Pfeiffer, R., and Gail, M.H., *Haplotype analysis in population genetics and association studies*. Pharmacogenomics, 2003. **4**(2): p. 171-8.
506. Levenstien, M.A., Ott, J., and Gordon, D., *Are molecular haplotypes worth the time and expense? A cost-effective method for applying molecular haplotypes*. PLoS Genet, 2006. **2**(8): p. e127.
507. Morris, R.W. and Kaplan, N.L., *On the advantage of haplotype analysis in the presence of multiple disease susceptibility alleles*. Genet Epidemiol, 2002. **23**(3): p. 221-33.
508. Clark, A.G., *The role of haplotypes in candidate gene studies*. Genet Epidemiol, 2004. **27**(4): p. 321-33.
509. Lorenz, A.J., Hamblin, M.T., and Jannink, J.L., *Performance of single nucleotide polymorphisms versus haplotypes for genome-wide association analysis in barley*. PLoS One, 2010. **5**(11): p. e14079.
510. Anderson, C.A., et al., *Data quality control in genetic case-control association studies*. Nat Protoc, 2010. **5**(9): p. 1564-73.
511. Kochi, Y., et al., *Genetics of rheumatoid arthritis: underlying evidence of ethnic differences*. J Autoimmun, 2009. **32**(3-4): p. 158-62.

512. Amos, W., Driscoll, E., and Hoffman, J.I., *Candidate genes versus genome-wide associations: which are better for detecting genetic susceptibility to infectious disease?* Proc Biol Sci, 2011. **278**(1709): p. 1183-8.
513. Evans, D.M. and Purcell, S., *Power calculations in genetic studies*. Cold Spring Harb Protoc, 2012. **2012**(6): p. 664-74.
514. Men, A.E., et al., *Sanger DNA Sequencing*, in *Next-Generation Genome Sequencing: Towards Personalized Medicine*, Janitz, M., Editor 2008, Wiley-Blackwell. p. 1-12.
515. Shendure, J. and Ji, H., *Next-generation DNA sequencing*. Nat Biotechnol, 2008. **26**(10): p. 1135-45.
516. Gilissen, C., et al., *Disease gene identification strategies for exome sequencing*. Eur J Hum Genet, 2012. **20**(5): p. 490-7.
517. Zhang, W., Cui, H., and Wong, L.J., *Application of Next Generation Sequencing to Molecular Diagnosis of Inherited Diseases*. Top Curr Chem, 2014. **336**: p. 19-45.
518. Clement, J.P.T., et al., *Association and stoichiometry of K(ATP) channel subunits*. Neuron, 1997. **18**(5): p. 827-38.
519. Gloyn, A.L., et al., *Activating mutations in the gene encoding the ATP-sensitive potassium-channel subunit Kir6.2 and permanent neonatal diabetes*. N Engl J Med, 2004. **350**(18): p. 1838-49.
520. Miki, T., et al., *ATP-sensitive K⁺ channels in the hypothalamus are essential for the maintenance of glucose homeostasis*. Nat Neurosci, 2001. **4**(5): p. 507-12.
521. Yamada, K., et al., *Protective role of ATP-sensitive potassium channels in hypoxia-induced generalized seizure*. Science, 2001. **292**(5521): p. 1543-6.
522. Gloyn, A.L., et al., *Large-scale association studies of variants in genes encoding the pancreatic beta-cell KATP channel subunits Kir6.2 (KCNJ11) and SUR1 (ABCC8) confirm that the KCNJ11 E23K variant is associated with type 2 diabetes*. Diabetes, 2003. **52**(2): p. 568-72.
523. Chistiakov, D.A., et al., *Genetic variations in the pancreatic ATP-sensitive potassium channel, beta-cell dysfunction, and susceptibility to type 2 diabetes*. Acta Diabetol, 2009. **46**(1): p. 43-9.
524. Thomas, P., Ye, Y., and Lightner, E., *Mutation of the pancreatic islet inward rectifier Kir6.2 also leads to familial persistent hyperinsulinemic hypoglycemia of infancy*. Hum Mol Genet, 1996. **5**(11): p. 1809-12.
525. Gloyn, A.L., et al., *KCNJ11 activating mutations are associated with developmental delay, epilepsy and neonatal diabetes syndrome and other neurological features*. Eur J Hum Genet, 2006. **14**(7): p. 824-30.
526. Sang, Y., et al., *AV59M KCNJ11 gene mutation leading to intermediate DEND syndrome in a Chinese child*. J Pediatr Endocrinol Metab, 2011. **24**(9-10): p. 763-6.
527. Proks, P., et al., *Molecular basis of Kir6.2 mutations associated with neonatal diabetes or neonatal diabetes plus neurological features*. Proc Natl Acad Sci USA, 2004. **101**(50): p. 17539-44.
528. Nestorowicz, A., et al., *A nonsense mutation in the inward rectifier potassium channel gene, Kir6.2, is associated with familial hyperinsulinism*. Diabetes, 1997. **46**(11): p. 1743-8.
529. Koster, J.C., et al., *Targeted overactivity of beta cell K(ATP) channels induces profound neonatal diabetes*. Cell, 2000. **100**(6): p. 645-54.
530. Shyng, S.L., et al., *Functional analyses of novel mutations in the sulfonylurea receptor 1 associated with persistent hyperinsulinemic hypoglycemia of infancy*. Diabetes, 1998. **47**(7): p. 1145-51.
531. Gloyn, A.L., et al., *Activating mutations in the gene encoding the ATP-sensitive potassium-channel subunit Kir6.2 and permanent neonatal diabetes*. N Eng J Med, 2004. **350**(18): p. 1838-49.

532. Harada, H., et al., *Phosphorylation and inactivation of BAD by mitochondria-anchored protein kinase A*. Mol Cell, 1999. **3**(4): p. 413-22.
533. Yang, E., et al., *Bad, a heterodimeric partner for Bcl-XL and Bcl-2, displaces Bax and promotes cell death*. Cell, 1995. **80**(2): p. 285-91.
534. Claveria, C., Martinez, A.C. and Torres, M., *A Bax/Bak-independent mitochondrial death pathway triggered by Drosophila Grim GH3 domain in mammalian cells*. J Biol Chem, 2004. **279**(2): p. 1368-75.
535. Niquet, J. and Wasterlain, C.G., *Bim, Bad, and Bax: a deadly combination in epileptic seizures*. J Clin Invest, 2004. **113**(7): p. 960-2.
536. Roy, S.S., et al., *Bad targets the permeability transition pore independent of Bax or Bak to switch between Ca²⁺-dependent cell survival and death*. Mol Cell, 2009. **33**(3): p. 377-88.
537. Zha, J., et al., *Serine phosphorylation of death agonist BAD in response to survival factor results in binding to 14-3-3 not BCL-X(L)*. Cell, 1996. **87**(4): p. 619-28.
538. Fletcher, J.I., et al., *Apoptosis is triggered when prosurvival Bcl-2 proteins cannot restrain Bax*. Proc Natl Acad Sci USA, 2008. **105**(47): p. 18081-7.
539. Chiang, C.W., et al., *Protein phosphatase 2A activates the proapoptotic function of BAD in interleukin-3-dependent lymphoid cells by a mechanism requiring 14-3-3 dissociation*. Blood, 2001. **97**(5): p. 1289-97.
540. Echeverry, R., et al., *The cytokine tumor necrosis factor-like weak inducer of apoptosis and its receptor fibroblast growth factor-inducible 14 have a neuroprotective effect in the central nervous system*. J Neuroinflammation, 2012. **9**: p. 45.
541. Ishrat, T., et al., *Progesterone is neuroprotective against ischemic brain injury through its effects on the phosphoinositide 3-kinase/protein kinase B signaling pathway*. Neuroscience, 2012. **210**: p. 442-50.
542. Danial, N.N., et al., *BAD and glucokinase reside in a mitochondrial complex that integrates glycolysis and apoptosis*. Nature, 2003. **424**(6951): p. 952-6.
543. Arden, C., S. Baltrusch, and L. Agius, *Glucokinase regulatory protein is associated with mitochondria in hepatocytes*. FEBS Letters, 2006. **580**(8): p. 2065-70.
544. Danial, N.N., et al., *Dual role of proapoptotic BAD in insulin secretion and beta cell survival*. Nat Med, 2008. **14**(2): p. 144-53.
545. Datta, S.R., et al., *14-3-3 proteins and survival kinases cooperate to inactivate BAD by BH3 domain phosphorylation*. Mol Cell, 2000. **6**(1): p. 41-51.
546. Baggio, L.L. and D.J. Drucker, *Biology of incretins: GLP-1 and GIP*. Gastroenterology, 2007. **132**(6): p. 2131-57.
547. Nichols, C.G., *KATP channels as molecular sensors of cellular metabolism*. Nature, 2006. **440**(7083): p. 470-6.
548. Henshall, D.C., et al., *Activation of Bcl-2-associated death protein and counter-response of Akt within cell populations during seizure-induced neuronal death*. J Neurosci, 2002. **22**(19): p. 8458-65.
549. Noh, H.S., et al., *Ketogenic Diet Protects the Hippocampus From Kainic Acid Toxicity by Inhibiting the Dissociation of Bad From 14-3-3*. Cell, 2006. **1836**: p. 1829-1836.
550. Noh, H.S., et al., *Ketogenic diet protects the hippocampus from kainic acid toxicity by inhibiting the dissociation of bad from 14-3-3*. J Neurosci Res, 2006. **84**(8): p. 1829-36.
551. Bost, F., et al., *The extracellular signal-regulated kinase isoform ERK1 is specifically required for in vitro and in vivo adipogenesis*. Diabetes, 2005. **54**(2): p. 402-11.
552. Shah, P., et al., *Short chain fatty acids induce TH gene expression via ERK-dependent phosphorylation of CREB protein*. Brain Res, 2006. **1107**(1): p. 13-23.

553. Domouzoglou, E.M. and Maratos-Flier, E., *Fibroblast growth factor 21 is a metabolic regulator that plays a role in the adaptation to ketosis*. Am J Clin Nutr, 2011. **93**(4): p. 901S-905S.
554. Andjelkovic, M., et al., *Activation and phosphorylation of a pleckstrin homology domain containing protein kinase (RAC-PK/PKB) promoted by serum and protein phosphatase inhibitors*. Proc Natl Acad Sci USA, 1996. **93**(12): p. 5699-704.
555. Liu, L. and Kumar, S., *Evolutionary balancing is critical for correctly forecasting disease-associated amino acid variants*. Mol Biol Evol, 2013. **30**(6): p. 1252-7.
556. Flanagan, S.E., et al., *Update of mutations in the genes encoding the pancreatic beta-cell K(ATP) channel subunits Kir6.2 (KCNJ11) and sulfonylurea receptor 1 (ABCC8) in diabetes mellitus and hyperinsulinism*. Hum Mutat, 2009. **30**(2): p. 170-80.
557. Inoue, H., et al., *Sequence variants in the pancreatic islet beta-cell inwardly rectifying K⁺ channel Kir6.2 (Bir) gene: identification and lack of role in Caucasian patients with NIDDM*. Diabetes, 1997. **46**(3): p. 502-7.
558. Purcell, S., et al., *PLINK: a tool set for whole-genome association and population-based linkage analyses*. Am J Hum Genet, 2007. **81**(3): p. 559-75.
559. Engels, W.R., *Exact tests for Hardy-Weinberg proportions*. Genetics, 2009. **183**(4): p. 1431-41.
560. Morris, A.P. and Zeggini, E., *An evaluation of statistical approaches to rare variant analysis in genetic association studies*. Genet Epidemiol, 2010. **34**(2): p. 188-93.
561. Clarke, G.M., et al., *Basic statistical analysis in genetic case-control studies*. Nat Protoc, 2011. **6**(2): p. 121-33.
562. Gao, X., *Multiple testing corrections for imputed SNPs*. Genet Epidemiol, 2011. **35**(3): p. 154-8.
563. Lunetta, K.L., *Statistical Primer for Cardiovascular Research*. Circulation, 2008. **118**: p. 96-101.
564. Knijnenburg, T.A., et al., *Fewer permutations, more accurate P-values*. Bioinformatics, 2009. **25**(12): p. i161-8.
565. Gao, X., Starmer, J., and Martin, E.R., *A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms*. Genet Epidemiol, 2008. **32**(4): p. 361-9.
566. Churchill, G.A. and Doerge, R.W., *Empirical threshold values for quantitative trait mapping*. Genetics, 1994. **138**(3): p. 963-71.
567. Chen, C., et al., *PICARA, an analytical pipeline providing probabilistic inference about a priori candidates genes underlying genome-wide association QTL in plants*. PLoS One, 2012. **7**(11): p. e46596.
568. Han, B., Kang, H.M., and Eskin, E., *Rapid and accurate multiple testing correction and power estimation for millions of correlated markers*. PLoS Genet, 2009. **5**(4): p. e1000456.
569. Cheung, C.L., et al., *Hip geometry variation is associated with bone mineralization pathway gene variants: The Framingham Study*. J Bone Miner Res, 2010. **25**(7): p. 1564-71.
570. De La Vega, F., Bustamante, C.D., and Leal, S.M., *GENOME-WIDE ASSOCIATION MAPPING AND RARE ALLELES: FROM POPULATION GENOMICS TO PERSONALIZED MEDICINE - Session Introduction*. Pac Symp Biocomput, 2011: p. 74-5.
571. Abo, R., et al., *Automated construction and testing of multi-locus gene-gene associations*. Bioinformatics, 2011. **27**(1): p. 134-6.
572. Babron, M.C., et al., *Rare and low frequency variant stratification in the UK population: description and impact on association tests*. PLoS One, 2012. **7**(10): p. e46519.

573. Wellcome Trust Case Control Consortium, *Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls*. *Nature*, 2007. **447**(7145): p. 661-78.
574. Chang, X. and Wang, K., *wANNOVAR: annotating genetic variants for personal genomes via the web*. *J Med Genet*, 2012. **49**(7): p. 433-6.
575. Ng, P.C. and Henikoff, S., *SIFT: Predicting amino acid changes that affect protein function*. *Nucleic Acids Res*, 2003. **31**(13): p. 3812-4.
576. Adzhubei, I.A., et al., *A method and server for predicting damaging missense mutations*. *Nat Methods*, 2010. **7**(4): p. 248-9.
577. Chun, S. and Fay, J.C., *Identification of deleterious mutations within three human genomes*. *Genome Res*, 2009. **19**(9): p. 1553-61.
578. Schwarz, J.M., et al., *MutationTaster evaluates disease-causing potential of sequence alterations*. *Nat Methods*, 2010. **7**(8): p. 575-6.
579. Davydov, E.V., et al., *Identifying a high fraction of the human genome to be under selective constraint using GERP++*. *PLoS Comput Biol*, 2010. **6**(12): p. e1001025.
580. Shastry, B.S., *SNPs: impact on gene function and phenotype*. *Methods Mol Biol*, 2009. **578**: p. 3-22.
581. Menashe, I., Rosenberg, P.S., and Chen, B.E., *PGA: power calculator for case-control genetic association analyses*. *BMC Genet*, 2008. **9**: p. 36.
582. Kasperaviciute, D., et al., *Common genetic variation and susceptibility to partial epilepsies: a genome-wide association study*. *Brain*, 2010. **133**(Pt 7): p. 2136-47.
583. Stenson, P.D., et al., *The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine*. *Hum Genet*, 2013.
584. Purcell, S., Cherny, S.S., and Sham, P.C., *Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits*. *Bioinformatics*, 2003. **19**(1): p. 149-50.
585. Thomas, D.C. and Witte, J.S., *Point: population stratification: a problem for case-control studies of candidate-gene associations?* *Cancer Epidemiol Biomarkers Prev*, 2002. **11**(6): p. 505-12.
586. Tabor, H.K., Risch, N.J., and Myers, R.M., *Candidate-gene approaches for studying complex genetic traits: practical considerations*. *Nat Rev Genet*, 2002. **3**(5): p. 391-7.
587. Cartegni, L., Chew, S.L., and Krainer, A.R., *Listening to silence and understanding nonsense: exonic mutations that affect splicing*. *Nat Rev Genet*, 2002. **3**(4): p. 285-98.
588. Tennessen, J.A., et al., *Evolution and functional impact of rare coding variation from deep sequencing of human exomes*. *Science*, 2012. **337**(6090): p. 64-9.
589. Babenko, A.P., et al., *Activating mutations in the ABCC8 gene in neonatal diabetes mellitus*. *N Engl J Med*, 2006. **355**(5): p. 456-66.
590. Chamary, J.V., Parmley, J.L., and Hurst, L.D., *Hearing silence: non-neutral evolution at synonymous sites in mammals*. *Nat Rev Genet*, 2006. **7**(2): p. 98-108.
591. Edwards, N.C., et al., *Characterization of coding synonymous and non-synonymous variants in ADAMTS13 using ex vivo and in silico approaches*. *PLoS One*, 2012. **7**(6): p. e38864.
592. Hunt, R., et al., *Silent (synonymous) SNPs: should we care about them?*. *Methods Mol Biol*, 2009. **578**: p. 23-39.
593. Komar, A.A., *A pause for thought along the co-translational folding pathway*. *Trends Biochem Sci*, 2009. **34**(1): p. 16-24.

594. Hansen, L., et al., *Amino acid polymorphisms in the ATP-regulatable inward rectifier Kir6.2 and their relationships to glucose- and tolbutamide-induced insulin secretion, the insulin sensitivity index, and NIDDM*. Diabetes, 1997. **46**(3): p. 508-12.
595. Riedel, M.J., Steckley, D.C., and Light, P.E., *Current status of the E23K Kir6.2 polymorphism: implications for type-2 diabetes*. Hum Genet, 2005. **116**(3): p. 133-45.
596. Schwanstecher, C., Meyer, U., and Schwanstecher, M., *K(IR)6.2 polymorphism predisposes to type 2 diabetes by inducing overactivity of pancreatic beta-cell ATP-sensitive K(+) channels*. Diabetes, 2002. **51**(3): p. 875-9.
597. Schwanstecher, C., et al., *The common single nucleotide polymorphism E23K in K(IR)6.2 sensitizes pancreatic beta-cell ATP-sensitive potassium channels toward activation through nucleoside diphosphates*. Diabetes, 2002. **51 Suppl 3**: p. S363-7.
598. Sakura, H., et al., *Sequence variations in the human Kir6.2 gene, a subunit of the beta-cell ATP-sensitive K-channel: no association with NIDDM in while Caucasian subjects or evidence of abnormal function when expressed in vitro*. Diabetologia, 1996. **39**(10): p. 1233-6.
599. Hani, E.H., et al., *Missense mutations in the pancreatic islet beta cell inwardly rectifying K+ channel gene (KIR6.2/BIR): a meta-analysis suggests a role in the polygenic basis of Type II diabetes mellitus in Caucasians*. Diabetologia, 1998. **41**(12): p. 1511-5.
600. Laukkanen, O., et al., *Polymorphisms of the SUR1 (ABCC8) and Kir6.2 (KCNJ11) genes predict the conversion from impaired glucose tolerance to type 2 diabetes. The Finnish Diabetes Prevention Study*. J Clin Endocrinol Metab, 2004. **89**(12): p. 6286-90.
601. Chauhan, G., et al., *Impact of common variants of PPARG, KCNJ11, TCF7L2, SLC30A8, HHEX, CDKN2A, IGF2BP2, and CDKAL1 on the risk of type 2 diabetes in 5,164 Indians*. Diabetes, 2010. **59**(8): p. 2068-74.
602. Omori, S., et al., *Association of CDKAL1, IGF2BP2, CDKN2A/B, HHEX, SLC30A8, and KCNJ11 with susceptibility to type 2 diabetes in a Japanese population*. Diabetes, 2008. **57**(3): p. 791-5.
603. Yang, L., et al., *Association between KCNJ11 gene polymorphisms and risk of type 2 diabetes mellitus in East Asian populations: a meta-analysis in 42,573 individuals*. Mol Biol Rep, 2012. **39**(1): p. 645-59.
604. Hu, C., et al., *PPARG, KCNJ11, CDKAL1, CDKN2A-CDKN2B, IDE-KIF11-HHEX, IGF2BP2 and SLC30A8 are associated with type 2 diabetes in a Chinese population*. PLoS One, 2009. **4**(10): p. e7643.
605. Zhou, D., et al., *The E23K variation in the KCNJ11 gene is associated with type 2 diabetes in Chinese and East Asian population*. J Hum Genet, 2009. **54**(7): p. 433-5.
606. Sanghera, D.K., et al., *Impact of nine common type 2 diabetes risk polymorphisms in Asian Indian Sikhs: PPARG2 (Pro12Ala), IGF2BP2, TCF7L2 and FTO variants confer a significant risk*. BMC Med Genet, 2008. **9**: p. 59.
607. Thorsby, P.M., et al., *Comparison of genetic risk in three candidate genes (TCF7L2, PPARG, KCNJ11) with traditional risk factors for type 2 diabetes in a population-based study--the HUNT study*. Scand J Clin Lab Invest, 2009. **69**(2): p. 282-7.
608. Gupta, V., et al., *A validation study of type 2 diabetes-related variants of the TCF7L2, HHEX, KCNJ11, and ADIPOQ genes in one endogamous ethnic group of north India*. Ann Hum Genet, 2010. **74**(4): p. 361-8.
609. Gong, B., et al., *The effect of KCNJ11 polymorphism on the risk of type 2 diabetes: a global meta-analysis based on 49 case-control studies*. DNA Cell Biol, 2012. **31**(5): p. 801-10.

610. Cukras, C.A., Jeliaskova, I., and Nichols, C.G., *The role of NH₂-terminal positive charges in the activity of inward rectifier KATP channels*. J Gen Physiol, 2002. **120**(3): p. 437-46.
611. Nielsen, E.M., et al., *The E23K variant of Kir6.2 associates with impaired post-OGTT serum insulin response and increased risk of type 2 diabetes*. Diabetes, 2003. **52**(2): p. 573-7.
612. Fischer, A., et al., *KCNJ11 E23K affects diabetes risk and is associated with the disposition index: results of two independent German cohorts*. Diabetes Care, 2008. **31**(1): p. 87-9.
613. 't Hart, L.M., et al., *Variations in insulin secretion in carriers of the E23K variant in the KIR6.2 subunit of the ATP-sensitive K(+) channel in the beta-cell*. Diabetes, 2002. **51**(10): p. 3135-8.
614. Tschritter, O., et al., *The prevalent Glu23Lys polymorphism in the potassium inward rectifier 6.2 (KIR6.2) gene is associated with impaired glucagon suppression in response to hyperglycemia*. Diabetes, 2002. **51**(9): p. 2854-60.
615. Wilcox, G., *Insulin and insulin resistance*. Clin Biochem Rev, 2005. **26**(2): p. 19-39.
616. Coventry, A., et al., *Deep resequencing reveals excess rare recent variants consistent with explosive population growth*. Nat Commun, 2010. **1**: p. 131.
617. Kwon, J.M. and Goate, A.M., *The candidate gene approach*. Alcohol Res Health, 2000. **24**(3): p. 164-8.
618. Lander, E.S., *The new genomics: global views of biology*. Science, 1996. **274**(5287): p. 536-9.
619. Collins, F.S., Guyer, M.S., and Charkravarti, A., *Variations on a theme: cataloging human DNA sequence variation*. Science, 1997. **278**(5343): p. 1580-1.
620. Reich, D.E. and Lander, E.S., *On the allelic spectrum of human disease*. Trends Genet, 2001. **17**(9): p. 502-10.
621. Li, M., Li, C., and Guan, W., *Evaluation of coverage variation of SNP chips for genome-wide association studies*. Eur J Hum Genet, 2008. **16**(5): p. 635-43.
622. Wilkening, S., et al., *Is there still a need for candidate gene approaches in the era of genome-wide association studies?* Genomics, 2009. **93**(5): p. 415-9.
623. Wooten, E.C. and Huggins, G.S., *Mind the dbGAP: the application of data mining to identify biological mechanisms*. Mol Interv, 2011. **11**(2): p. 95-102.
624. Teo, Y.Y., *Common statistical issues in genome-wide association studies: a review on power, data quality control, genotype calling and population structure*. Curr Opin Lipidol, 2008. **19**(2): p. 133-43.
625. Neale, B.M. and Purcell, S., *The positives, protocols, and perils of genome-wide association*. Am J Med Genet B Neuropsychiatr Genet, 2008. **147B**(7): p. 1288-94.
626. Teo, Y.Y., et al., *On the usage of HWE for identifying genotyping errors*. Ann Hum Genet, 2007. **71**(Pt 5): p. 701-3; author reply 704.
627. Need, A.C., et al., *A genome-wide investigation of SNPs and CNVs in schizophrenia*. PLoS Genet, 2009. **5**(2): p. e1000373.
628. Pharm, T.D., *Computational Biology: Issues and Applications in Oncology* 2009, New York, USA: Springer Science+Business Media.
629. Devlin, B. and Roeder, K., *Genomic control for association studies*. Biometrics, 1999. **55**(4): p. 997-1004.
630. Reich, D.E. and Goldstein, D.B., *Detecting association in a case-control study while correcting for population stratification*. Genet Epidemiol, 2001. **20**(1): p. 4-16.
631. Price, A.L., et al., *New approaches to population stratification in genome-wide association studies*. Nat Rev Genet, 2010. **11**(7): p. 459-63.

632. Li, Q. and Yu, K., *Improved correction for population stratification in genome-wide association studies by identifying hidden population structures*. Genet Epidemiol, 2008. **32**(3): p. 215-26.
633. Stevens, E.L., et al., *Inference of relationships in population data using identity-by-descent and identity-by-state*. PLoS Genet, 2011. **7**(9): p. e1002287.
634. Luan, T., et al., *The importance of identity-by-state information for the accuracy of genomic selection*. Genet Sel Evol, 2012. **44**: p. 28.
635. Pritchard, J.K., Stephens, M., and Donnelly, P., *Inference of population structure using multilocus genotype data*. Genetics, 2000. **155**(2): p. 945-59.
636. Adeyemo, A., et al., *A genome-wide association study of hypertension and blood pressure in African Americans*. PLoS Genet, 2009. **5**(7): p. e1000564.
637. Paterson, A.D., et al., *A genome-wide association study identifies a novel major locus for glycemic control in type 1 diabetes, as measured by both A1C and glucose*. Diabetes, 2010. **59**(2): p. 539-49.
638. Raska, P., et al., *European American stratification in ovarian cancer case control data: the utility of genome-wide data for inferring ancestry*. PLoS One, 2012. **7**(5): p. e35235.
639. Tian, C., Gregersen, P.K., and Seldin, M.F., *Accounting for ancestry: population substructure and genome-wide association studies*. Hum Mol Genet, 2008. **17**(R2): p. R143-50.
640. Johnson, R.C., et al., *Accounting for multiple comparisons in a genome-wide association study (GWAS)*. BMC Genomics, 2010. **11**: p. 724.
641. Cheverud, J.M., *A simple correction for multiple comparisons in interval mapping genome scans*. Heredity (Edinb), 2001. **87**(Pt 1): p. 52-8.
642. Nyholt, D.R., *A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other*. Am J Hum Genet, 2004. **74**(4): p. 765-9.
643. Moskvina, V. and Schmidt, K.M., *On multiple-testing correction in genome-wide association studies*. Genet Epidemiol, 2008. **32**(6): p. 567-73.
644. Storey, J.D. and Tibshirani, R., *Statistical significance for genomewide studies*. Proc Natl Acad Sci U S A, 2003. **100**(16): p. 9440-5.
645. Hochberg, Y. and Benjamini, Y., *More powerful procedures for multiple significance testing*. Stat Med, 1990. **9**(7): p. 811-8.
646. Bennett, C.M., Wolford, G.L., and Miller, M.B., *The principled control of false positives in neuroimaging*. Soc Cogn Affect Neurosci, 2009. **4**(4): p. 417-22.
647. Hoffman, G.E., Logsdon, B.A., and Mezey, J.G., *PUMA: a unified framework for penalized multiple regression analysis of GWAS data*. PLoS Comput Biol, 2013. **9**(6): p. e1003101.
648. Benjamini, Y. and Hochberg, Y., *Controlling the false discovery rate: a practical and powerful approach to multiple testing*. Journal of the Royal Statistical Society. , 1995. **57**(1): p. 289-300.
649. Pennisi, E., *The Biology of Genomes. Disease risk links to gene regulation*. Science, 2011. **332**(6033): p. 1031.
650. Kumar, V., Wijmenga, C., and Withoff, S., *From genome-wide association studies to disease mechanisms: celiac disease as a model for autoimmune diseases*. Semin Immunopathol, 2012. **34**(4): p. 567-80.
651. Klein, R.J., et al., *Complement factor H polymorphism in age-related macular degeneration*. Science, 2005. **308**(5720): p. 385-9.
652. Barber, M.J., et al., *Genome-wide association of lipid-lowering response to statins in combined study populations*. PLoS One, 2010. **5**(3): p. e9763.

653. Wittkowski, K.M., et al., *From single-SNP to wide-locus: genome-wide association studies identifying functionally related genes and intragenic regions in small sample studies*. Pharmacogenomics, 2013. **14**(4): p. 391-401.
654. Lambert, C.G. and Black, L.J., *Learning from our GWAS mistakes: from experimental design to scientific method*. Biostatistics, 2012. **13**(2): p. 195-203.
655. Cantor, R.M., Lange, K., and Sinsheimer, J.S., *Prioritizing GWAS results: A review of statistical methods and recommendations for their application*. Am J Hum Genet, 2010. **86**(1): p. 6-22.
656. Ioannidis, J.P., et al., *A road map for efficient and reliable human genome epidemiology*. Nat Genet, 2006. **38**(1): p. 3-5.
657. de Bakker, P.I., et al., *Practical aspects of imputation-driven meta-analysis of genome-wide association studies*. Hum Mol Genet, 2008. **17**(R2): p. R122-8.
658. Marchini, J., Donnelly, P., and Cardon, L.R., *Genome-wide strategies for detecting multiple loci that influence complex diseases*. Nat Genet, 2005. **37**(4): p. 413-7.
659. Emmert-Streib, F. and Glazko, G.V., *Pathway analysis of expression data: deciphering functional building blocks of complex diseases*. PLoS Comput Biol, 2011. **7**(5): p. e1002053.
660. Neale, B.M., et al., *Testing for an unusual distribution of rare variants*. PLoS Genet, 2011. **7**(3): p. e1001322.
661. Peng, G., et al., *Gene and pathway-based second-wave analysis of genome-wide association studies*. Eur J Hum Genet, 2010. **18**(1): p. 111-7.
662. Hong, M.G., et al., *Strategies and issues in the detection of pathway enrichment in genome-wide association studies*. Hum Genet, 2009. **126**(2): p. 289-301.
663. Lee, P.H., et al., *INRICH: interval-based enrichment analysis for genome-wide association studies*. Bioinformatics, 2012. **28**(13): p. 1797-9.
664. Franke, L., et al., *Reconstruction of a functional human gene network, with an application for prioritizing positional candidate genes*. Am J Hum Genet, 2006. **78**(6): p. 1011-25.
665. Jia, P., Liu, Y., and Zhao, Z., *Integrative pathway analysis of genome-wide association studies and gene expression data in prostate cancer*. BMC Syst Biol, 2012. **6 Suppl 3**: p. S13.
666. Wu, G. and Zhi, D., *Pathway-based approaches for sequencing-based genome-wide association studies*. Genet Epidemiol, 2013. **37**(5): p. 478-94.
667. Wang, K., Li, M., and Bucan, M., *Pathway-based approaches for analysis of genomewide association studies*. Am J Hum Genet, 2007. **81**(6): p. 1278-83.
668. Farber, C.R., *Systems-level analysis of genome-wide association data*. G3 (Bethesda), 2013. **3**(1): p. 119-29.
669. Holden, M., et al., *GSEA-SNP: applying gene set enrichment analysis to SNP data from genome-wide association studies*. Bioinformatics, 2008. **24**(23): p. 2784-5.
670. Dennis, G., Jr., et al., *DAVID: Database for Annotation, Visualization, and Integrated Discovery*. Genome Biol, 2003. **4**(5): p. P3.
671. Mi, H., Muruganujan, A., and Thomas, P.D., *PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees*. Nucleic Acids Res, 2013. **41**(Database issue): p. D377-86.
672. O'Dushlaine, C., et al., *The SNP ratio test: pathway analysis of genome-wide association datasets*. Bioinformatics, 2009. **25**(20): p. 2762-3.
673. Mukherjee, S., et al., *Gene-based GWAS and biological pathway analysis of the resilience of executive functioning*. Brain Imaging Behav, 2013.
674. Law, B., et al., *Effects of population structure and admixture on exact tests for association between Loci*. Genetics, 2003. **164**(1): p. 381-7.

675. Visscher, P.M., et al., *Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings*. PLoS Genet, 2006. **2**(3): p. e41.
676. Pongpanich, M., Sullivan, P.F., and Tzeng, J.Y., *A quality control algorithm for filtering SNPs in genome-wide association studies*. Bioinformatics, 2010. **26**(14): p. 1731-7.
677. Slager, S.L., et al., *Common variation at 6p21.31 (BAK1) influences the risk of chronic lymphocytic leukemia*. Blood, 2012. **120**(4): p. 843-6.
678. Wigginton, J.E., Cutler, D.J., and Abecasis, G.R., *A note on exact tests of Hardy-Weinberg equilibrium*. Am J Hum Genet, 2005. **76**(5): p. 887-93.
679. Unoki, H., et al., *SNPs in KCNQ1 are associated with susceptibility to type 2 diabetes in East Asian and European populations*. Nat Genet, 2008. **40**(9): p. 1098-102.
680. Favret, S., et al., *Deficiency in the metabolite receptor SUCNR1 (GPR91) leads to outer retinal lesions*. Aging (Albany NY), 2013. **5**(6): p. 427-44.
681. Newton-Cheh, C., et al., *Common variants at ten loci influence QT interval duration in the QTGEN Study*. Nat Genet, 2009. **41**(4): p. 399-406.
682. SanGiovanni, J.P., et al., *DNA sequence variants in PPARGC1A, a gene encoding a coactivator of the omega-3 LCPUFA sensing PPAR-RXR transcription complex, are associated with NV AMD and AMD-associated loci in genes of complement and VEGF signaling pathways*. PLoS One, 2013. **8**(1): p. e53155.
683. Cox, D.G. and Kraft, P., *Quantification of the power of Hardy-Weinberg equilibrium testing to detect genotyping error*. Hum Hered, 2006. **61**(1): p. 10-4.
684. Hosking, L., et al., *Detection of genotyping errors by Hardy-Weinberg equilibrium testing*. Eur J Hum Genet, 2004. **12**(5): p. 395-9.
685. Leal, S.M., *Detection of genotyping errors and pseudo-SNPs via deviations from Hardy-Weinberg equilibrium*. Genet Epidemiol, 2005. **29**(3): p. 204-14.
686. Abdellaoui, A., et al., *Population structure, migration, and diversifying selection in the Netherlands*. Eur J Hum Genet, 2013. **21**(11): p. 1277-85.
687. Price, A.L., et al., *Long-range LD can confound genome scans in admixed populations*. Am J Hum Genet, 2008. **83**(1): p. 132-5; author reply 135-9.
688. Wang, G. and Leal, S.M., *Genome-wide Association Study using GenABEL*. 2012 04/01/2014].
689. McCarthy, M.I., et al., *Genome-wide association studies for complex traits: consensus, uncertainty and challenges*. Nat Rev Genet, 2008. **9**(5): p. 356-69.
690. Ridley, J., et al., *An unexpected influence of widely used significance thresholds on the distribution of reported P-values*. J Evol Biol, 2007. **20**(3): p. 1082-9.
691. Salyakina, D., et al., *Evaluation of Nyholt's procedure for multiple testing correction*. Hum Hered, 2005. **60**(1): p. 19-25; discussion 61-2.
692. Sidak, Z., *Rectangular confidence regions for the means of multivariate normal distributions*. J Am Stat Assoc, 1967. **62**(31): p. 626-633.
693. Wright, F.A., et al., *Genome-wide association and linkage identify modifier loci of lung disease severity in cystic fibrosis at 11p13 and 20q13.2*. Nat Genet, 2011. **43**(6): p. 539-46.
694. Simino, J., et al., *Variants on chromosome 6p22.3 associated with blood pressure in the HyperGEN study: follow-up of FBPP quantitative trait loci*. Am J Hypertens, 2011. **24**(11): p. 1227-33.
695. Barrett, J.C., et al., *Haploview: analysis and visualization of LD and haplotype maps*. Bioinformatics, 2005. **21**(2): p. 263-5.
696. Pruim, R.J., et al., *LocusZoom: regional visualization of genome-wide association scan results*. Bioinformatics, 2010. **26**(18): p. 2336-7.
697. Seelow, D., Schwarz, J.M., and Schuelke, M., *GeneDistiller--distilling candidate genes from linkage intervals*. PLoS One, 2008. **3**(12): p. e3874.

698. Becker, K.G., et al., *The genetic association database*. Nat Genet, 2004. **36**(5): p. 431-432.
699. Ramos, E.M., et al., *Phenotype-Genotype Integrator (PheGenI): synthesizing genome-wide association study (GWAS) data with existing genomic resources*. Eur J Hum Genet, 2014. **22**(1): p. 144-7.
700. Gui, H., et al., *Comparisons of seven algorithms for pathway analysis using the WTCCC Crohn's Disease dataset*. BMC Res Notes, 2011. **4**: p. 386.
701. Huang da, W., Sherman, B.T., and Lempicki, R.A., *Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources*. Nat Protoc, 2009. **4**(1): p. 44-57.
702. Elbers, C.C., et al., *Using genome-wide pathway analysis to unravel the etiology of complex diseases*. Genet Epidemiol, 2009. **33**(5): p. 419-31.
703. Mulligan, P., et al., *CDYL bridges REST and histone methyltransferases for gene repression and suppression of cellular transformation*. Mol Cell, 2008. **32**(5): p. 718-26.
704. Mabley, J.G., et al., *Inhibition of poly (ADP-ribose) synthetase by gene disruption or inhibition with 5-iodo-6-amino-1,2-benzopyrone protects mice from multiple-low-dose-streptozotocin-induced diabetes*. Br J Pharmacol, 2001. **133**(6): p. 909-19.
705. Masutani, M., et al., *Poly(ADP-ribose) polymerase gene disruption conferred mice resistant to streptozotocin-induced diabetes*. Proc Natl Acad Sci U S A, 1999. **96**(5): p. 2301-4.
706. Klimov, E., et al., *Genomic organisation and tissue specific expression of ABLIM2 gene in human, mouse and rat*. Biochim Biophys Acta, 2005. **1730**(1): p. 1-9.
707. Moore, S., et al., *Distinct regulatory mechanisms act to establish and maintain Pax3 expression in the developing neural tube*. PLoS Genet, 2013. **9**(10): p. e1003811.
708. Kapranov, P., Willingham, A.T., and Gingeras, T.R., *Genome-wide transcription and the implications for genomic organization*. Nat Rev Genet, 2007. **8**(6): p. 413-23.
709. Butter, F., et al., *Proteome-wide analysis of disease-associated SNPs that show allele-specific transcription factor binding*. PLoS Genet, 2012. **8**(9): p. e1002982.
710. Kumar, V., et al., *Human disease-associated genetic variation impacts large intergenic non-coding RNA expression*. PLoS Genet, 2013. **9**(1): p. e1003201.
711. Mercer, T.R., Dinger, M.E., and Mattick, J.S., *Long non-coding RNAs: insights into functions*. Nat Rev Genet, 2009. **10**(3): p. 155-9.
712. Kathiresan, S., et al., *A genome-wide association study for blood lipid phenotypes in the Framingham Heart Study*. BMC Med Genet, 2007. **8 Suppl 1**: p. S17.
713. Meigs, J.B., et al., *Genome-wide association with diabetes-related traits in the Framingham Heart Study*. BMC Med Genet, 2007. **8 Suppl 1**: p. S16.
714. Haiman, C.A., et al., *Genome-wide testing of putative functional exonic variants in relationship with breast and prostate cancer risk in a multiethnic population*. PLoS Genet, 2013. **9**(3): p. e1003419.
715. Stranger, B.E., Stahl, E.A., and Raj, T., *Progress and promise of genome-wide association studies for human complex trait genetics*. Genetics, 2011. **187**(2): p. 367-83.
716. Visscher, P.M., et al., *Evidence-based psychiatric genetics, AKA the false dichotomy between common and rare variant hypotheses*. Mol Psychiatry, 2012. **17**(5): p. 474-85.
717. Kleinjan, D.A. and van Heyningen, V., *Long-range control of gene expression: emerging mechanisms and disruption in disease*. Am J Hum Genet, 2005. **76**(1): p. 8-32.
718. Wu, M.C., et al., *Powerful SNP-set analysis for case-control genome-wide association studies*. Am J Hum Genet, 2010. **86**(6): p. 929-42.

719. McDonald, W.I., et al., *Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the diagnosis of multiple sclerosis*. Ann Neurol, 2001. **50**(1): p. 121-7.
720. Ioannidis, J.P., Thomas, G., and Daly, M.J., *Validating, augmenting and refining genome-wide association signals*. Nat Rev Genet, 2009. **10**(5): p. 318-29.
721. London, S.J., *Gene-air pollution interactions in asthma*. Proc Am Thorac Soc, 2007. **4**(3): p. 217-20.
722. Pirinen, M., Donnelly, P., and Spencer, C.C., *Including known covariates can reduce power to detect genetic effects in case-control studies*. Nat Genet, 2012. **44**(8): p. 848-51.
723. Sasieni, P.D., *From genotypes to genes: doubling the sample size*. Biometrics, 1997. **53**(4): p. 1253-61.
724. Miclaus, K., et al., *Batch effects in the BRLMM genotype calling algorithm influence GWAS results for the Affymetrix 500K array*. Pharmacogenomics J, 2010. **10**(4): p. 336-46.
725. Jakubek, Y.A. and Cutler, D.J., *A model of binding on DNA microarrays: understanding the combined effect of probe synthesis failure, cross-hybridization, DNA fragmentation and other experimental details of affymetrix arrays*. BMC Genomics, 2012. **13**: p. 737.
726. Chacon-Cortes, D., et al., *Comparison of genomic DNA extraction techniques from whole blood samples: a time, cost and quality evaluation study*. Mol Biol Rep, 2012. **39**(5): p. 5961-6.
727. Neel, J.V., *Diabetes mellitus: a "thrifty" genotype rendered detrimental by "progress"?* Am J Hum Genet, 1962. **14**: p. 353-62.
728. Hegele, R.A., *Genes and environment in type 2 diabetes and atherosclerosis in aboriginal Canadians*. Curr Atheroscler Rep, 2001. **3**(3): p. 216-21.
729. Story, M., et al., *The epidemic of obesity in American Indian communities and the need for childhood obesity-prevention programs*. Am J Clin Nutr, 1999. **69**(4 Suppl): p. 747S-754S.
730. Marchini, J., et al., *The effects of human population structure on large genetic association studies*. Nat Genet, 2004. **36**(5): p. 512-7.
731. Jagtap, P. and Szabo, C., *Poly(ADP-ribose) polymerase and the therapeutic effects of its inhibitors*. Nat Rev Drug Discov, 2005. **4**(5): p. 421-40.
732. Virag, L., *Structure and function of poly(ADP-ribose) polymerase-1: role in oxidative stress-related pathologies*. Curr Vasc Pharmacol, 2005. **3**(3): p. 209-14.
733. Burkart, V., et al., *Mice lacking the poly(ADP-ribose) polymerase gene are resistant to pancreatic beta-cell destruction and diabetes development induced by streptozocin*. Nat Med, 1999. **5**(3): p. 314-9.
734. Szabo, C., et al., *Poly(ADP-Ribose) polymerase is activated in subjects at risk of developing type 2 diabetes and is associated with impaired vascular reactivity*. Circulation, 2002. **106**(21): p. 2680-6.
735. Garcia Soriano, F., et al., *Diabetic endothelial dysfunction: the role of poly(ADP-ribose) polymerase activation*. Nat Med, 2001. **7**(1): p. 108-13.
736. Obrosova, I.G. and Julius, U.A., *Role for poly(ADP-ribose) polymerase activation in diabetic nephropathy, neuropathy and retinopathy*. Curr Vasc Pharmacol, 2005. **3**(3): p. 267-83.
737. Obrosova, I.G., et al., *PARP inhibition or gene deficiency counteracts intraepidermal nerve fiber loss and neuropathic pain in advanced diabetic neuropathy*. Free Radic Biol Med, 2008. **44**(6): p. 972-81.
738. Macgregor, S., et al., *Genome-wide association study identifies a new melanoma susceptibility locus at 1q21.3*. Nat Genet, 2011. **43**(11): p. 1114-8.

739. Liu, H.P., et al., *Evaluation of the poly(ADP-ribose) polymerase-1 gene variants in Alzheimer's disease*. J Clin Lab Anal, 2010. **24**(3): p. 182-6.
740. Infante, J., et al., *Poly (ADP-ribose) polymerase-1 (PARP-1) genetic variants are protective against Parkinson's disease*. J Neurol Sci, 2007. **256**(1-2): p. 68-70.
741. Brighina, L., et al., *Association analysis of PARP1 polymorphisms with Parkinson's disease*. Parkinsonism Relat Disord, 2011. **17**(9): p. 701-4.
742. Egi, Y., et al., *Neuroprotective effects of a novel water-soluble poly(ADP-ribose) polymerase-1 inhibitor, MP-124, in in vitro and in vivo models of cerebral ischemia*. Brain Res, 2011. **1389**: p. 169-76.
743. Takahashi, K., et al., *Post-treatment with an inhibitor of poly(ADP-ribose) polymerase attenuates cerebral damage in focal ischemia*. Brain Res, 1999. **829**(1-2): p. 46-54.
744. Moroni, F. and Chiarugi, A., *Post-ischemic brain damage: targeting PARP-1 within the ischemic neurovascular units as a realistic avenue to stroke treatment*. FEBS J, 2009. **276**(1): p. 36-45.
745. Iwashita, A., et al., *Neuroprotective effects of a novel poly(ADP-ribose) polymerase-1 inhibitor, 2-[3-[4-(4-chlorophenyl)-1-piperazinyl] propyl]-4(3H)-quinazolinone (FR255595), in an in vitro model of cell death and in mouse 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine model of Parkinson's disease*. J Pharmacol Exp Ther, 2004. **309**(3): p. 1067-78.
746. Meier, H.L., et al., *Benzamide, a poly(ADP-ribose) polymerase inhibitor, is neuroprotective against soman-induced seizure-related brain damage*. Ann N Y Acad Sci, 1999. **890**: p. 330-5.
747. Virag, L. and Szabo, C., *The therapeutic potential of poly(ADP-ribose) polymerase inhibitors*. Pharmacol Rev, 2002. **54**(3): p. 375-429.
748. Kim, M.Y., Zhang, T., and Kraus, W.L., *Poly(ADP-ribosyl)ation by PARP-1: 'PAR-laying' NAD⁺ into a nuclear signal*. Genes Dev, 2005. **19**(17): p. 1951-67.
749. Gonzalez, C., et al., *Unexpected sensitivity of nonobese diabetic mice with a disrupted poly(ADP-Ribose) polymerase-1 gene to streptozotocin-induced and spontaneous diabetes*. Diabetes, 2002. **51**(5): p. 1470-6.
750. Barrientos, T., et al., *Two novel members of the ABLIM protein family, ABLIM-2 and -3, associate with STARS and directly bind F-actin*. J Biol Chem, 2007. **282**(11): p. 8393-403.
751. Lundquist, E.A., et al., *UNC-115, a conserved protein with predicted LIM and actin-binding domains, mediates axon guidance in C. elegans*. Neuron, 1998. **21**(2): p. 385-92.
752. Struckhoff, E.C. and Lundquist, E.A., *The actin-binding protein UNC-115 is an effector of Rac signaling during axon pathfinding in C. elegans*. Development, 2003. **130**(4): p. 693-704.
753. Erkman, L., et al., *A POU domain transcription factor-dependent program regulates axon pathfinding in the vertebrate visual system*. Neuron, 2000. **28**(3): p. 779-92.
754. Baldwin, C.T., et al., *Mutations in PAX3 associated with Waardenburg syndrome type I*. Hum Mutat, 1994. **3**(3): p. 205-11.
755. Hoth, C.F., et al., *Mutations in the paired domain of the human PAX3 gene cause Klein-Waardenburg syndrome (WS-III) as well as Waardenburg syndrome type I (WS-I)*. Am J Hum Genet, 1993. **52**(3): p. 455-62.
756. Begum, S., et al., *Cell-type-specific regulation of distinct sets of gene targets by Pax3 and Pax3/FKHR*. Oncogene, 2005. **24**(11): p. 1860-72.
757. Joven, A., et al., *Spatiotemporal patterns of Pax3, Pax6, and Pax7 expression in the developing brain of a urodele amphibian, Pleurodeles waltl*. J Comp Neurol, 2013. **521**(17): p. 3913-53.

758. Liem, K.F., Jr., et al., *Dorsal differentiation of neural plate cells induced by BMP-mediated signals from epidermal ectoderm*. Cell, 1995. **82**(6): p. 969-79.
759. Bang, A.G., et al., *Expression of Pax-3 in the lateral neural plate is dependent on a Wnt-mediated signal from posterior nonaxial mesoderm*. Dev Biol, 1999. **212**(2): p. 366-80.
760. Litingtung, Y. and Chiang, C., *Specification of ventral neuron types is mediated by an antagonistic interaction between Shh and Gli3*. Nat Neurosci, 2000. **3**(10): p. 979-85.
761. Ratnapriya, R., et al., *A locus for juvenile myoclonic epilepsy maps to 2q33-q36*. Hum Genet, 2010. **128**(2): p. 123-30.
762. Chen, J., et al., *Clinical significance and prognostic value of PAX3 expression in human glioma*. J Mol Neurosci, 2012. **47**(1): p. 52-8.
763. Grupe, A., et al., *A scan of chromosome 10 identifies a novel locus showing strong association with late-onset Alzheimer disease*. Am J Hum Genet, 2006. **78**(1): p. 78-88.
764. Funato, Y. and Miki, H., *Reversible oxidation of PRL family protein-tyrosine phosphatases*. Methods, 2013.
765. Zeng, Q., et al., *PRL-3 and PRL-1 promote cell migration, invasion, and metastasis*. Cancer Res, 2003. **63**(11): p. 2716-22.
766. Matter, W.F., et al., *Role of PRL-3, a human muscle-specific tyrosine phosphatase, in angiotensin-II signaling*. Biochem Biophys Res Commun, 2001. **283**(5): p. 1061-8.
767. Schroeder, M.D., Symowicz, J., and Schuler, L.A., *PRL modulates cell cycle regulators in mammary tumor epithelial cells*. Mol Endocrinol, 2002. **16**(1): p. 45-57.
768. Takano, S., et al., *PRL-1, a protein tyrosine phosphatase, is expressed in neurons and oligodendrocytes in the brain and induced in the cerebral cortex following transient forebrain ischemia*. Brain Res Mol Brain Res, 1996. **40**(1): p. 105-15.
769. Zeng, Q., Hong, W., and Tan, Y.H., *Mouse PRL-2 and PRL-3, two potentially prenylated protein tyrosine phosphatases homologous to PRL-1*. Biochem Biophys Res Commun, 1998. **244**(2): p. 421-7.
770. Bachelot, A. and Binart, N., *Reproductive role of prolactin*. Reproduction, 2007. **133**(2): p. 361-9.
771. Al-Sakkaf, K.A., et al., *Possible role for protein kinase B in the anti-apoptotic effect of prolactin in rat Nb2 lymphoma cells*. J Endocrinol, 2000. **167**(1): p. 85-92.
772. Bishop, J.D., et al., *Prolactin activates mammalian target-of-rapamycin through phosphatidylinositol 3-kinase and stimulates phosphorylation of p70S6K and 4E-binding protein-1 in lymphoma cells*. J Endocrinol, 2006. **190**(2): p. 307-12.
773. Peng, L., et al., *PRL-3 promotes the motility, invasion, and metastasis of LoVo colon cancer cells through PRL-3-integrin beta1-ERK1/2 and-MMP2 signaling*. Mol Cancer, 2009. **8**: p. 110.
774. Fiordalisi, J.J., Keller, P.J., and Cox, A.D., *PRL tyrosine phosphatases regulate rho family GTPases to promote invasion and motility*. Cancer Res, 2006. **66**(6): p. 3153-61.
775. Dumauval, C.M., et al., *Integrated analysis of global mRNA and protein expression data in HEK293 cells overexpressing PRL-1*. PLoS One, 2013. **8**(9): p. e72977.
776. Liang, F., et al., *PRL3 promotes cell invasion and proliferation by down-regulation of Csk leading to Src activation*. J Biol Chem, 2007. **282**(8): p. 5413-9.
777. Neilson, L.M., et al., *Coactivation of janus tyrosine kinase (Jak)1 positively modulates prolactin-Jak2 signaling in breast cancer: recruitment of ERK and signal transducer and activator of transcription (Stat)3 and enhancement of Akt and Stat5a/b pathways*. Mol Endocrinol, 2007. **21**(9): p. 2218-32.

778. Gall, C.M. and Lynch, G., *Integrins, Synaptic Plasticity and Epileptogenesis*, in *Recent Advances in Epilepsy Research*, D.K. Binder and Scharfman, H.E., Editors. 2004, Kluwer Academic / Plenum Publishers. p. 12-33.
779. Bonni, A., et al., *Cell survival promoted by the Ras-MAPK signaling pathway by transcription-dependent and -independent mechanisms*. *Science*, 1999. **286**(5443): p. 1358-62.
780. Zheng, Y.L., et al., *Cdk5 Modulation of mitogen-activated protein kinase signaling regulates neuronal survival*. *Mol Biol Cell*, 2007. **18**(2): p. 404-13.
781. Kitauro, J., et al., *Dysregulation of Src family kinases in mast cells from epilepsy-resistant ASK versus epilepsy-prone EL mice*. *J Immunol*, 2007. **178**(1): p. 455-62.
782. Salter, M.W. and Kalia, L.V., *Src kinases: a hub for NMDA receptor regulation*. *Nat Rev Neurosci*, 2004. **5**(4): p. 317-28.
783. Rocha, L. and Cavalheiro, E.A., *Pharmacoresistance in Epilepsy: From Genes and Molecules to Promising Therapies*. 2013: Springer.
784. Plas, D.R., Rathmell, J.C., and Thompson, C.B., *Homeostatic control of lymphocyte survival: potential origins and implications*. *Nat Immunol*, 2002. **3**(6): p. 515-21.
785. Riley, J.K., et al., *Phosphatidylinositol 3-kinase activity is critical for glucose metabolism and embryo survival in murine blastocysts*. *J Biol Chem*, 2006. **281**(9): p. 6010-9.
786. Marko, A.J., et al., *Induction of glucose metabolism in stimulated T lymphocytes is regulated by mitogen-activated protein kinase signaling*. *PLoS One*, 2010. **5**(11): p. e15425.
787. Rosenzweig, T., et al., *Src tyrosine kinase regulates insulin-induced activation of protein kinase C (PKC) delta in skeletal muscle*. *Cell Signal*, 2004. **16**(11): p. 1299-308.
788. Ueda, S., et al., *Crucial role of the small GTPase Rac1 in insulin-stimulated translocation of glucose transporter 4 to the mouse skeletal muscle sarcolemma*. *FASEB J*, 2010. **24**(7): p. 2254-61.
789. Ueda, S., Kataoka, T., and Satoh, T., *Activation of the small GTPase Rac1 by a specific guanine-nucleotide-exchange factor suffices to induce glucose uptake into skeletal-muscle cells*. *Biol Cell*, 2008. **100**(11): p. 645-57.
790. Sylow, L., et al., *Rac1 signaling is required for insulin-stimulated glucose uptake and is dysregulated in insulin-resistant murine and human skeletal muscle*. *Diabetes*, 2013. **62**(6): p. 1865-75.
791. Gustav Smith, J. and Newton-Cheh, C., *Genome-Wide Association Study in Humans*, in *Cardiovascular Genomics, Methods in Molecular Biology*, K. DiPetrillo, Editor 2009, Humana Press. p. 231-257.
792. Marchini, J. and Howie, B., *Genotype imputation for genome-wide association studies*. *Nat Rev Genet*, 2010. **11**(7): p. 499-511.
793. Lu, J.T., et al., *Characterizing linkage disequilibrium and evaluating imputation power of human genomic insertion-deletion polymorphisms*. *Genome Biol*, 2012. **13**(2): p. R15.
794. Pantanowitz, A. and Marwala, T., *Evaluating the Impact of Missing Data Imputation*, in *Advanced Data Mining and Applications*, R. Huang, et al., Editors. 2009, Springer. p. 577-586.
795. Editorial, *On beyond GWAS*. *Nat Genet*, 2010. **42**(7): p. 551.
796. Gorlov, I.P., et al., *Shifting paradigm of association studies: value of rare single-nucleotide polymorphisms*. *Am J Hum Genet*, 2008. **82**(1): p. 100-12.
797. Hunt, K.A., et al., *Negligible impact of rare autoimmune-locus coding-region variants on missing heritability*. *Nature*, 2013. **498**(7453): p. 232-5.

798. Nelson, M.R., et al., *An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people*. Science, 2012. **337**(6090): p. 100-4.
799. Daye, Z.J., Li, H., and Wei, Z., *A powerful test for multiple rare variants association studies that incorporates sequencing qualities*. Nucleic Acids Res, 2012. **40**(8): p. e60.
800. Bloom, J.S., et al., *Finding the sources of missing heritability in a yeast cross*. Nature, 2013. **494**(7436): p. 234-7.
801. Hugot, J.P., et al., *Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease*. Nature, 2001. **411**(6837): p. 599-603.
802. Ogura, Y., et al., *A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease*. Nature, 2001. **411**(6837): p. 603-6.
803. Rivas, M.A., et al., *Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease*. Nat Genet, 2011. **43**(11): p. 1066-73.
804. Bonnefond, A., et al., *Rare MTNR1B variants impairing melatonin receptor 1B function contribute to type 2 diabetes*. Nat Genet, 2012. **44**(3): p. 297-301.
805. Holm, H., et al., *A rare variant in MYH6 is associated with high risk of sick sinus syndrome*. Nat Genet, 2011. **43**(4): p. 316-20.
806. Wei, P., Liu, X., and Fu, Y.X., *Incorporating predicted functions of nonsynonymous variants into gene-based analysis of exome sequencing data: a comparative study*. BMC Proc, 2011. **5 Suppl 9**: p. S20.
807. Raychaudhuri, S., *Mapping rare and common causal alleles for complex human diseases*. Cell, 2011. **147**(1): p. 57-69.
808. Gibson, G., *Rare and common variants: twenty arguments*. Nat Rev Genet, 2011. **13**(2): p. 135-45.
809. Rommens, J.M., et al., *Identification of the cystic fibrosis gene: chromosome walking and jumping*. Science, 1989. **245**(4922): p. 1059-65.
810. Easton, D.F., et al., *A systematic genetic assessment of 1,433 sequence variants of unknown clinical significance in the BRCA1 and BRCA2 breast cancer-predisposition genes*. Am J Hum Genet, 2007. **81**(5): p. 873-83.
811. Hughes, A.L., *Small effective population sizes and rare nonsynonymous variants in potyviruses*. Virology, 2009. **393**(1): p. 127-34.
812. Ng, P.C. and Henikoff, S., *Predicting the effects of amino acid substitutions on protein function*. Annu Rev Genomics Hum Genet, 2006. **7**: p. 61-80.
813. Stenson, P.D., et al., *The Human Gene Mutation Database: 2008 update*. Genome Med, 2009. **1**(1): p. 13.
814. Abecasis, G.R., et al., *An integrated map of genetic variation from 1,092 human genomes*. Nature, 2012. **491**(7422): p. 56-65.
815. MacArthur, D.G., et al., *A systematic survey of loss-of-function variants in human protein-coding genes*. Science, 2012. **335**(6070): p. 823-8.
816. Chang, Y.F., Imam, J.S., and Wilkinson, M.F., *The nonsense-mediated decay RNA surveillance pathway*. Annu Rev Biochem, 2007. **76**: p. 51-74.
817. Woolfe, A., Mullikin, J.C., and Elnitski, L., *Genomic features defining exonic variants that modulate splicing*. Genome Biol, 2010. **11**(2): p. R20.
818. Bisio, A., et al., *Functional analysis of CDKN2A/p16INK4a 5'-UTR variants predisposing to melanoma*. Hum Mol Genet, 2010. **19**(8): p. 1479-91.
819. MacArthur, D.G. and Tyler-Smith, C., *Loss-of-function variants in the genomes of healthy humans*. Hum Mol Genet, 2010. **19**(R2): p. R125-30.
820. Saleh, M., et al., *Differential modulation of endotoxin responsiveness by human caspase-12 polymorphisms*. Nature, 2004. **429**(6987): p. 75-9.

821. Keren, H., Lev-Maor, G., and Ast, G., *Alternative splicing and evolution: diversification, exon definition and function*. Nat Rev Genet, 2010. **11**(5): p. 345-55.
822. Cirak, S., et al., *Exon skipping and dystrophin restoration in patients with Duchenne muscular dystrophy after systemic phosphorodiamidate morpholino oligomer treatment: an open-label, phase 2, dose-escalation study*. Lancet, 2011. **378**(9791): p. 595-605.
823. Williams, L.E. and Wernegreen, J.J., *Sequence context of indel mutations and their effect on protein evolution in a bacterial endosymbiont*. Genome Biol Evol, 2013. **5**(3): p. 599-605.
824. Kryukov, G.V., Pennacchio, L.A., and Sunyaev, S.R., *Most rare missense alleles are deleterious in humans: implications for complex disease and association studies*. Am J Hum Genet, 2007. **80**(4): p. 727-39.
825. Mullaney, J.M., et al., *Small insertions and deletions (INDELs) in human genomes*. Hum Mol Genet, 2010. **19**(R2): p. R131-6.
826. Leushkin, E.V., Bazykin, G.A., and Kondrashov, A.S., *Strong mutational bias toward deletions in the Drosophila melanogaster genome is compensated by selection*. Genome Biol Evol, 2013. **5**(3): p. 514-24.
827. McCarroll, S.A., et al., *Deletion polymorphism upstream of IRGM associated with altered IRGM expression and Crohn's disease*. Nat Genet, 2008. **40**(9): p. 1107-12.
828. Venables, J.P., *Downstream intronic splicing enhancers*. FEBS Lett, 2007. **581**(22): p. 4127-31.
829. Lee, Y., et al., *Variants affecting exon skipping contribute to complex traits*. PLoS Genet, 2012. **8**(10): p. e1002998.
830. Lorson, C.L., et al., *A single nucleotide in the SMN gene regulates splicing and is responsible for spinal muscular atrophy*. Proc Natl Acad Sci U S A, 1999. **96**(11): p. 6307-11.
831. Chatterjee, S. and Pal, J.K., *Role of 5'- and 3'-untranslated regions of mRNAs in human diseases*. Biol Cell, 2009. **101**(5): p. 251-62.
832. Pickering, B.M. and Willis, A.E., *The implications of structured 5' untranslated regions on translation and disease*. Semin Cell Dev Biol, 2005. **16**(1): p. 39-47.
833. Steinberg, M.H., et al., *Disorders of Hemoglobin: Genetics, Pathophysiology, and Clinical Management*. 2nd ed. 2009, USA: Cambridge University Press.
834. Reamon-Buettner, S.M., Cho, S.H., and Borlak, J., *Mutations in the 3'-untranslated region of GATA4 as molecular hotspots for congenital heart disease (CHD)*. BMC Med Genet, 2007. **8**: p. 38.
835. Hesketh, J., *3'-Untranslated regions are important in mRNA localization and translation: lessons from selenium and metallothionein*. Biochem Soc Trans, 2004. **32**(Pt 6): p. 990-3.
836. Kumar, P., Henikoff, S., and Ng, P.C., *Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm*. Nat Protoc, 2009. **4**(7): p. 1073-81.
837. Castellana, S. and Mazza, T., *Congruency in the prediction of pathogenic missense mutations: state-of-the-art web-based tools*. Brief Bioinform, 2013. **14**(4): p. 448-59.
838. Wei, Q., et al., *Testing computational prediction of missense mutation phenotypes: functional characterization of 204 mutations of human cystathionine beta synthase*. Proteins, 2010. **78**(9): p. 2058-74.
839. Siepel, A., et al., *Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes*. Genome Res, 2005. **15**(8): p. 1034-50.
840. Lopes, M.C., et al., *A combined functional annotation score for non-synonymous variants*. Hum Hered, 2012. **73**(1): p. 47-51.

841. Kumar, S., et al., *Evolutionary diagnosis method for variants in personal exomes*. Nat Methods, 2012. **9**(9): p. 855-6.
842. Lehmann, K.V. and Chen, T., *Exploring functional variant discovery in non-coding regions with SlnBaD*. Nucleic Acids Res, 2013. **41**(1): p. e7.
843. Chan, P.A., et al., *Interpreting missense variants: comparing computational methods in human disease genes CDKN2A, MLH1, MSH2, MECP2, and tyrosinase (TYR)*. Hum Mutat, 2007. **28**(7): p. 683-93.
844. Wu, J. and R. Jiang, *Prediction of deleterious nonsynonymous single-nucleotide polymorphism for human diseases*. ScientificWorldJournal, 2013. **2013**: p. 675851.
845. Liu, X., Jian, X., and Boerwinkle, E., *dbNSFP v2.0: A Database of Human Non-synonymous SNVs and Their Functional Predictions and Annotations*. Hum Mutat, 2013. **34**(9): p. E2393-402.
846. Wang, K., Li, M., and Hakonarson, H., *ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data*. Nucleic Acids Res, 2010. **38**(16): p. e164.
847. Liu, X., Jian, X., and Boerwinkle, E., *dbNSFP: a lightweight database of human nonsynonymous SNPs and their functional predictions*. Hum Mutat, 2011. **32**(8): p. 894-9.
848. Hicks, S., et al., *Prediction of missense mutation functionality depends on both the algorithm and sequence alignment employed*. Hum Mutat, 2011. **32**(6): p. 661-8.
849. Karchin, R., et al., *Classifying Variants of Undetermined Significance in BRCA2 with protein likelihood ratios*. Cancer Inform, 2008. **6**: p. 203-16.
850. Karchin, R., *Next generation tools for the annotation of human SNPs*. Brief Bioinform, 2009. **10**(1): p. 35-52.
851. Gonzalez-Perez, A. and Lopez-Bigas, N., *Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel*. Am J Hum Genet, 2011. **88**(4): p. 440-9.
852. Frousios, K., et al., *Predicting the functional consequences of non-synonymous DNA sequence variants - evaluation of bioinformatics tools and development of a consensus strategy*. Genomics, 2013. **102**(4): p. 223-8.
853. Li, M.X., et al., *Predicting mendelian disease-causing non-synonymous single nucleotide variants in exome sequencing studies*. PLoS Genet, 2013. **9**(1): p. e1003143.
854. Bodmer, W. and Bonilla, C., *Common and rare variants in multifactorial susceptibility to common diseases*. Nat Genet, 2008. **40**(6): p. 695-701.
855. Emond, M.J., et al., *Exome sequencing of extreme phenotypes identifies DCTN4 as a modifier of chronic Pseudomonas aeruginosa infection in cystic fibrosis*. Nat Genet, 2012. **44**(8): p. 886-9.
856. Yilmaz, Y.E. and Bull, S.B., *Are quantitative trait-dependent sampling designs cost-effective for analysis of rare and common variants?* BMC Proc, 2011. **5 Suppl 9**: p. S111.
857. Cohen, J.C., et al., *Multiple rare alleles contribute to low plasma levels of HDL cholesterol*. Science, 2004. **305**(5685): p. 869-72.
858. Huang, B.E. and Lin, D.Y., *Efficient association mapping of quantitative trait loci with selective genotyping*. Am J Hum Genet, 2007. **80**(3): p. 567-76.
859. Herbert, A., et al., *A common genetic variant is associated with adult and childhood obesity*. Science, 2006. **312**(5771): p. 279-83.
860. Lanktree, M.B., et al., *Extremes of unexplained variation as a phenotype: an efficient approach for genome-wide association studies of cardiovascular disease*. Circ Cardiovasc Genet, 2010. **3**(2): p. 215-21.

861. Li, D., et al., *Using extreme phenotype sampling to identify the rare causal variants of quantitative traits in association studies*. Genet Epidemiol, 2011. **35**(8): p. 790-9.
862. Liu, D.J. and Leal, S.M., *A unified framework for detecting rare variant quantitative trait associations in pedigree and unrelated individuals via sequence data*. Hum Hered, 2012. **73**(2): p. 105-22.
863. Cirulli, E.T. and Goldstein, D.B., *Uncovering the roles of rare variants in common disease through whole-genome sequencing*. Nat Rev Genet, 2010. **11**(6): p. 415-25.
864. Gallais, A., Moreau, L., and Charcosset, A., *Detection of marker-QTL associations by studying change in marker frequencies with selection*. Theor Appl Genet, 2007. **114**(4): p. 669-81.
865. Sen, S., Johannes, F., and Broman, K.W., *Selective genotyping and phenotyping strategies in a complex trait context*. Genetics, 2009. **181**(4): p. 1613-26.
866. Huyghe, J.R., et al., *Exome array analysis identifies new loci and low-frequency variants influencing insulin processing and secretion*. Nat Genet, 2013. **45**(2): p. 197-201.
867. Panagiotou, O.A., Evangelou, E., and Ioannidis, J.P., *Genome-wide significant associations for variants with minor allele frequency of 5% or less--an overview: A HuGE review*. Am J Epidemiol, 2010. **172**(8): p. 869-89.
868. Zheng, H.F., et al., *Effect of genome-wide genotyping and reference panels on rare variants imputation*. J Genet Genomics, 2012. **39**(10): p. 545-50.
869. Aschard, H., et al., *Combining effects from rare and common genetic variants in an exome-wide association study of sequence data*. BMC Proc, 2011. **5 Suppl 9**: p. S44.
870. Manolio, T.A., *Bringing genome-wide association findings into clinical use*. Nat Rev Genet, 2013. **14**(8): p. 549-58.
871. Lefevre, J.H., et al., *Role of rare variants in undetermined multiple adenomatous polyposis and early-onset colorectal cancer*. J Hum Genet, 2012. **57**(11): p. 709-16.
872. Shi, G. and Rao, D.C., *Optimum designs for next-generation sequencing to discover rare variants for common complex disease*. Genet Epidemiol, 2011. **35**(6): p. 572-9.
873. Chen, M.Y.M., *Regulation and Function of Non-Exonic Recursive Splicing*, in Department of Biological Sciences 2011, Carnegie Mellon University: Pittsburgh.
874. Ng, S.B., et al., *Targeted capture and massively parallel sequencing of 12 human exomes*. Nature, 2009. **461**(7261): p. 272-6.
875. Myers, R.A., et al., *A population genetic approach to mapping neurological disorder genes using deep resequencing*. PLoS Genet, 2011. **7**(2): p. e1001318.
876. Coon, H., et al., *Possible association between autism and variants in the brain-expressed tryptophan hydroxylase gene (TPH2)*. Am J Med Genet B Neuropsychiatr Genet, 2005. **135B**(1): p. 42-6.
877. Walitza, S., et al., *Transmission disequilibrium of polymorphic variants in the tryptophan hydroxylase-2 gene in attention-deficit/hyperactivity disorder*. Mol Psychiatry, 2005. **10**(12): p. 1126-32.
878. Mossner, R., et al., *Transmission disequilibrium of polymorphic variants in the tryptophan hydroxylase-2 gene in children and adolescents with obsessive-compulsive disorder*. Int J Neuropsychopharmacol, 2006. **9**(4): p. 437-42.
879. Zill, P., et al., *SNP and haplotype analysis of a novel tryptophan hydroxylase isoform (TPH2) gene provide evidence for association with major depression*. Mol Psychiatry, 2004. **9**(11): p. 1030-6.
880. Zhang, X., et al., *Loss-of-function mutation in tryptophan hydroxylase-2 identified in unipolar major depression*. Neuron, 2005. **45**(1): p. 11-6.
881. De Luca, V., et al., *Promoter polymorphism of second tryptophan hydroxylase isoform (TPH2) in schizophrenia and suicidality*. Psychiatry Res, 2005. **134**(2): p. 195-8.

882. Zill, P., et al., *Single nucleotide polymorphism and haplotype analysis of a novel tryptophan hydroxylase isoform (TPH2) gene in suicide victims*. Biol Psychiatry, 2004. **56**(8): p. 581-6.
883. Liao, Y., et al., *SCN2A mutation associated with neonatal epilepsy, late-onset episodic ataxia, myoclonus, and pain*. Neurology, 2010. **75**(16): p. 1454-8.
884. Lauxmann, S., et al., *An SCN2A mutation in a family with infantile seizures from Madagascar reveals an increased subthreshold Na(+) current*. Epilepsia, 2013. **54**(9): p. e117-21.
885. Vecchi, M., et al., *Infantile epilepsy associated with mosaic 2q24 duplication including SCN2A and SCN3A*. Seizure, 2011. **20**(10): p. 813-6.
886. Cummings, A.C., et al., *Genome-wide association and linkage study in the Amish detects a novel candidate late-onset Alzheimer disease gene*. Ann Hum Genet, 2012. **76**(5): p. 342-51.
887. Lin, D.Y., D. Zeng, and Z.Z. Tang, *Quantitative trait analysis in sequencing studies under trait-dependent sampling*. Proc Natl Acad Sci USA, 2013. **110**(30): p. 12247-52.
888. Lynch, M., *Rate, molecular spectrum, and consequences of human mutation*. Proc Natl Acad Sci USA, 2010. **107**(3): p. 961-8.
889. Fernandez, J. and Caballero, A., *Accumulation of deleterious mutations and equalization of parental contributions in the conservation of genetic resources*. Heredity (Edinb), 2001. **86**(Pt 4): p. 480-8.
890. Young, C.D., et al., *Modulation of glucose transporter 1 (GLUT1) expression levels alters mouse mammary tumor cell growth in vitro and in vivo*. PLoS One, 2011. **6**(8): p. e23205.
891. Vannucci, S.J., Maher, F., and Simpson, I.A., *Glucose transporter proteins in brain: delivery of glucose to neurons and glia*. Glia, 1997. **21**(1): p. 2-21.
892. Petrovski, S., et al., *Genic intolerance to functional variation and the interpretation of personal genomes*. PLoS Genet, 2013. **9**(8): p. e1003709.
893. Ryu, S.W., et al., *Fas-associated factor 1, FAF1, is a member of Fas death-inducing signaling complex*. J Biol Chem, 2003. **278**(26): p. 24003-10.
894. Park, M.Y., et al., *Fas-associated factor-1 inhibits nuclear factor-kappaB (NF-kappaB) activity by interfering with nuclear translocation of the RelA (p65) subunit of NF-kappaB*. J Biol Chem, 2004. **279**(4): p. 2544-9.
895. Betarbet, R., et al., *Fas-associated factor 1 and Parkinson's disease*. Neurobiol Dis, 2008. **31**(3): p. 309-15.
896. Sul, J.W., et al., *Accumulation of the parkin substrate, FAF1, plays a key role in the dopaminergic neurodegeneration*. Hum Mol Genet, 2013. **22**(8): p. 1558-73.
897. Gill, R., et al., *Blocking brain-derived neurotrophic factor inhibits injury-induced hyperexcitability of hippocampal CA3 neurons*. Eur J Neurosci, 2013. **38**(11): p. 3554-66.
898. Murray, P.S. and Holmes, P.V., *An overview of brain-derived neurotrophic factor and implications for excitotoxic vulnerability in the hippocampus*. Int J Pept, 2011. **2011**: p. 654085.
899. Peralta, E.R., Martin, B.C., and Edinger, A.L., *Differential effects of TBC1D15 and mammalian Vps39 on Rab7 activation state, lysosomal morphology, and growth factor dependence*. J Biol Chem, 2010. **285**(22): p. 16814-21.
900. Zhang, X.M., et al., *TBC domain family, member 15 is a novel mammalian Rab GTPase-activating protein with substrate preference for Rab7*. Biochem Biophys Res Commun, 2005. **335**(1): p. 154-61.
901. Napolioni, V., *Recent patents on epilepsy genetics*. Recent Pat DNA Gene Seq, 2009. **3**(3): p. 183-92.

902. Ito, M., et al., *Phenotypes and genotypes in epilepsy with febrile seizures plus*. Epilepsy Res, 2006. **70 Suppl 1**: p. S199-205.
903. Kamiya, K., et al., *A nonsense mutation of the sodium channel gene SCN2A in a patient with intractable epilepsy and mental decline*. J Neurosci, 2004. **24**(11): p. 2690-8.
904. Ito, M., et al., *Seizure phenotypes of a family with missense mutations in SCN2A*. Pediatr Neurol, 2004. **31**(2): p. 150-2.
905. Nguyen, T., et al., *Interactions of the human MCM-BP protein with MCM complex components and Dbf4*. PLoS One, 2012. **7**(4): p. e35931.
906. Ishimi, Y., et al., *Inhibition of Mcm4,6,7 helicase activity by phosphorylation with cyclin A/Cdk2*. J Biol Chem, 2000. **275**(21): p. 16235-41.
907. Chibazakura, T., et al., *Cyclin A promotes S-phase entry via interaction with the replication licensing factor Mcm7*. Mol Cell Biol, 2011. **31**(2): p. 248-55.
908. Ding, L. and Forsburg, S.L., *Schizosaccharomyces pombe minichromosome maintenance-binding protein (MCM-BP) antagonizes MCM helicase*. J Biol Chem, 2011. **286**(38): p. 32918-30.
909. Hubbi, M.E., et al., *MCM proteins are negative regulators of hypoxia-inducible factor 1*. Mol Cell, 2011. **42**(5): p. 700-12.
910. Graler, M.H., Bernhardt, G., and Lipp, M., *EDG6, a novel G-protein-coupled receptor related to receptors for bioactive lysophospholipids, is specifically expressed in lymphoid tissue*. Genomics, 1998. **53**(2): p. 164-9.
911. Graler, M.H., et al., *The sphingosine 1-phosphate receptor S1P4 regulates cell shape and motility via coupling to Gi and G12/13*. J Cell Biochem, 2003. **89**(3): p. 507-19.
912. Van Brocklyn, J.R., et al., *Sphingosine-1-phosphate is a ligand for the G protein-coupled receptor EDG-6*. Blood, 2000. **95**(8): p. 2624-9.
913. Guerrero, M., et al., *Identification of a Novel Agonist of the Sphingosine 1-phosphate Receptor 4 (S1P4)*, in *Probe Reports from the NIH Molecular Libraries Program*2010: Bethesda (MD).
914. Toman, R.E. and Spiegel, S., *Lysophospholipid receptors in the nervous system*. Neurochem Res, 2002. **27**(7-8): p. 619-27.
915. Zhang, W. and Liu, H.T., *MAPK signal pathways in the regulation of cell proliferation in mammalian cells*. Cell Res, 2002. **12**(1): p. 9-18.
916. Li, T., et al., *SH2D4A regulates cell proliferation via the ERalpha/PLC-gamma/PKC pathway*. BMB Rep, 2009. **42**(8): p. 516-22.
917. Lu, Z. and Xu, S., *ERK1/2 MAP kinases in cell survival and apoptosis*. IUBMB Life, 2006. **58**(11): p. 621-31.
918. Bae, S.S., et al., *Proteolytic cleavage of phospholipase C-gamma1 during apoptosis in Molt-4 cells*. FASEB J, 2000. **14**(9): p. 1083-92.
919. Giachello, C.N., et al., *MAPK/Erk-dependent phosphorylation of synapsin mediates formation of functional synapses and short-term homosynaptic plasticity*. J Cell Sci, 2010. **123**(Pt 6): p. 881-93.
920. Yoshii, A. and Constantine-Paton, M., *Postsynaptic BDNF-TrkB signaling in synapse maturation, plasticity, and disease*. Dev Neurobiol, 2010. **70**(5): p. 304-22.
921. Van Epps-Fung, M., et al., *A role for phospholipase C activity in GLUT4-mediated glucose transport*. Endocrinology, 1997. **138**(12): p. 5170-5.
922. Kayali, A.G., et al., *Association of the insulin receptor with phospholipase C-gamma (PLCgamma) in 3T3-L1 adipocytes suggests a role for PLCgamma in metabolic signaling by insulin*. J Biol Chem, 1998. **273**(22): p. 13808-18.
923. Wright, D.C., et al., *The effects of phospholipase C inhibition on insulin-stimulated glucose transport in skeletal muscle*. Metabolism, 2002. **51**(3): p. 271-3.

924. Zhang, W., et al., *MAPK/ERK signaling regulates insulin sensitivity to control glucose metabolism in Drosophila*. PLoS Genet, 2011. **7**(12): p. e1002429.
925. Kohno, T., et al., *Sphingosine 1-phosphate promotes cell migration through the activation of Cdc42 in Edg-6/S1P4-expressing cells*. Genes Cells, 2003. **8**(8): p. 685-97.
926. Thomas, A., Giesler, T., and White, E., *p53 mediates bcl-2 phosphorylation and apoptosis via activation of the Cdc42/JNK1 pathway*. Oncogene, 2000. **19**(46): p. 5259-69.
927. Zhao, J., et al., *Down-regulation Cdc42 attenuates neuronal apoptosis through inhibiting MLK3/JNK3 cascade during ischemic reperfusion in rat hippocampus*. Cell Signal, 2007. **19**(4): p. 831-43.
928. Wang, Z., E. Oh, and Thurmond, D.C., *Glucose-stimulated Cdc42 signaling is essential for the second phase of insulin secretion*. J Biol Chem, 2007. **282**(13): p. 9536-46.
929. Usui, I., et al., *Cdc42 is a Rho GTPase family member that can mediate insulin signaling to glucose transport in 3T3-L1 adipocytes*. J Biol Chem, 2003. **278**(16): p. 13765-74.
930. Li, B. and Leal, S.M., *Discovery of rare variants via sequencing: implications for the design of complex trait association studies*. PLoS Genet, 2009. **5**(5): p. e1000481.
931. Manolio, T.A., et al., *Finding the missing heritability of complex diseases*. Nature, 2009. **461**(7265): p. 747-53.
932. Eichler, E.E., et al., *Missing heritability and strategies for finding the underlying causes of complex disease*. Nat Rev Genet, 2010. **11**(6): p. 446-50.
933. Botstein, D. and Risch, N., *Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease*. Nat Genet, 2003. **33 Suppl**: p. 228-37.
934. Margulies, M., et al., *Genome sequencing in microfabricated high-density picolitre reactors*. Nature, 2005. **437**(7057): p. 376-80.
935. Bamshad, M.J., et al., *Exome sequencing as a tool for Mendelian disease gene discovery*. Nat Rev Genet, 2011. **12**(11): p. 745-55.
936. Ku, C.S., et al., *Exome sequencing: dual role as a discovery and diagnostic tool*. Ann Neurol, 2012. **71**(1): p. 5-14.
937. Snyder, M., Du, J., and Gerstein, M., *Personal genome sequencing: current approaches and challenges*. Genes Dev, 2010. **24**(5): p. 423-31.
938. Singleton, A., *Exome sequencing: making hay while the sun shines*. Lancet Neurol, 2011. **10**(10): p. 942-946.
939. Anderson, M.W. and Schrijver, I., *Next Generation DNA Sequencing and the Future of Genomic Medicine*. Genes, 2010. **1**(1): p. 38-69.
940. Choi, M., et al., *Genetic diagnosis by whole exome capture and massively parallel DNA sequencing*. Proc Natl Acad Sci USA, 2009. **106**(45): p. 19096-101.
941. Ng, S.B., et al., *Exome sequencing identifies the cause of a mendelian disorder*. Nat Genet, 2010. **42**(1): p. 30-5.
942. Walsh, T., et al., *Whole exome sequencing and homozygosity mapping identify mutation in the cell polarity protein GPSM2 as the cause of nonsyndromic hearing loss DFNB82*. Am J Hum Genet, 2010. **87**(1): p. 90-4.
943. Gilissen, C., et al., *Unlocking Mendelian disease using exome sequencing*. Genome Biol, 2011. **12**(9): p. 228.
944. Kiezun, A., et al., *Exome sequencing and the genetic basis of complex traits*. Nat Genet, 2012. **44**(6): p. 623-30.

945. Do, R., Kathiresan, S., and Abecasis, G.R., *Exome sequencing and complex disease: practical aspects of rare variant association studies*. Hum Mol Genet, 2012. **21**(R1): p. R1-9.
946. Musunuru, K., et al., *Exome sequencing, ANGPTL3 mutations, and familial combined hypolipidemia*. N Engl J Med, 2010. **363**(23): p. 2220-7.
947. Li, Y., et al., *Resequencing of 200 human exomes identifies an excess of low-frequency non-synonymous coding variants*. Nat Genet, 2010. **42**(11): p. 969-72.
948. Gazave, E., et al., *Population Growth Inflates the Per-Individual Number of Deleterious Mutations and Reduces Their Mean Effect*. Genetics, 2013. **195**(3): p. 969-78.
949. Marth, G.T., et al., *The functional spectrum of low-frequency coding variation*. Genome Biol, 2011. **12**(9): p. R84.
950. Feuk, L., et al., *Structural variants: changing the landscape of chromosomes and design of disease studies*. Hum Mol Genet, 2006. **15 Spec No 1**: p. R57-66.
951. Redon, R., et al., *Global variation in copy number in the human genome*. Nature, 2006. **444**(7118): p. 444-54.
952. Audrezet, M.P., et al., *Genomic rearrangements in the CFTR gene: extensive allelic heterogeneity and diverse mutational mechanisms*. Hum Mutat, 2004. **23**(4): p. 343-57.
953. Conrad, D.F., et al., *Origins and functional impact of copy number variation in the human genome*. Nature, 2010. **464**(7289): p. 704-12.
954. Mills, R.E., et al., *Mapping copy number variation by population-scale genome sequencing*. Nature, 2011. **470**(7332): p. 59-65.
955. Pinto, D., et al., *Functional impact of global rare copy number variation in autism spectrum disorders*. Nature, 2010. **466**(7304): p. 368-72.
956. Sebat, J., et al., *Strong association of de novo copy number mutations with autism*. Science, 2007. **316**(5823): p. 445-9.
957. Stefansson, H., et al., *Large recurrent microdeletions associated with schizophrenia*. Nature, 2008. **455**(7210): p. 232-6.
958. McCarthy, S.E., et al., *Microduplications of 16p11.2 are associated with schizophrenia*. Nat Genet, 2009. **41**(11): p. 1223-7.
959. Craddock, N., et al., *Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls*. Nature, 2010. **464**(7289): p. 713-20.
960. Stranger, B.E., et al., *Relative impact of nucleotide and copy number variation on gene expression phenotypes*. Science, 2007. **315**(5813): p. 848-53.
961. O'Roak, B.J., et al., *Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations*. Nat Genet, 2011. **43**(6): p. 585-9.
962. Panoutsopoulou, K., Tachmazidou, I., and Zeggini, E., *In search of low-frequency and rare variants affecting complex traits*. Hum Mol Genet, 2013. **22**(R1): p. R16-21.
963. Wagner, M.J., *Rare-variant genome-wide association studies: a new frontier in genetic analysis of complex traits*. Pharmacogenomics, 2013. **14**(4): p. 413-24.
964. Diogo, D., et al., *Rare, low-frequency, and common variants in the protein-coding sequence of biological candidate genes from GWASs contribute to risk of rheumatoid arthritis*. Am J Hum Genet, 2013. **92**(1): p. 15-27.
965. Aird, D., et al., *Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries*. Genome Biol, 2011. **12**(2): p. R18.
966. Pabinger, S., et al., *A survey of tools for variant analysis of next-generation genome sequencing data*. Brief Bioinform, 2013. First published online: January 21, 2013. doi: 10.1093/bib/bbs086.

967. Nielsen, R., et al., *Genotype and SNP calling from next-generation sequencing data*. Nat Rev Genet, 2011. **12**(6): p. 443-51.
968. Minoche, A.E., Dohm, J.C., and Himmelbauer, H., *Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and genome analyzer systems*. Genome Biol, 2011. **12**(11): p. R112.
969. Liu, Q., et al., *Steps to ensure accuracy in genotype and SNP calling from Illumina sequencing data*. BMC Genomics, 2012. **13 Suppl 8**: p. S8.
970. Dohm, J.C., et al., *Substantial biases in ultra-short read data sets from high-throughput DNA sequencing*. Nucleic Acids Res, 2008. **36**(16): p. e105.
971. Qu, W., Hashimoto, S., and Morishita, S., *Efficient frequency-based de novo short-read clustering for error trimming in next-generation sequencing*. Genome Res, 2009. **19**(7): p. 1309-15.
972. Li, H., Ruan, J., and Durbin, R., *Mapping short DNA sequencing reads and calling variants using mapping quality scores*. Genome Res, 2008. **18**(11): p. 1851-8.
973. Lunter, G. and Goodson, M., *Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads*. Genome Res, 2011. **21**(6): p. 936-9.
974. Langmead, B., et al., *Ultrafast and memory-efficient alignment of short DNA sequences to the human genome*. Genome Biol, 2009. **10**(3): p. R25.
975. Li, R., et al., *SOAP2: an improved ultrafast tool for short read alignment*. Bioinformatics, 2009. **25**(15): p. 1966-7.
976. Flicek, P. and Birney, E., *Sense from sequence reads: methods for alignment and assembly*. Nat Methods, 2009. **6**(11 Suppl): p. S6-S12.
977. Ruffalo, M., LaFramboise, T., and Koyuturk, M., *Comparative analysis of algorithms for next-generation sequencing read alignment*. Bioinformatics, 2011. **27**(20): p. 2790-6.
978. Illumina Inc. *An Introduction to Next-Generation Sequencing Technology*. res.illumina.com/documents/.../illumina_sequencing_introduction.pdf Accessed Online 24/09/13.
979. Liu, X., et al., *Variant callers for next-generation sequencing data: a comparison study*. PLoS One, 2013. **8**(9): p. e75619.
980. DePristo, M.A., et al., *A framework for variation discovery and genotyping using next-generation DNA sequencing data*. Nat Genet, 2011. **43**(5): p. 491-8.
981. Li, H., et al., *The Sequence Alignment/Map format and SAMtools*. Bioinformatics, 2009. **25**(16): p. 2078-9.
982. Zerbino, D.R. and Birney, E., *Velvet: algorithms for de novo short read assembly using de Bruijn graphs*. Genome Res, 2008. **18**(5): p. 821-9.
983. Dohm, J.C., et al., *SHARCGS, a fast and highly accurate short-read assembly algorithm for de novo genomic sequencing*. Genome Res, 2007. **17**(11): p. 1697-706.
984. Peng, Y., et al., *IDBA-tran: a more robust de novo de Bruijn graph assembler for transcriptomes with uneven expression levels*. Bioinformatics, 2013. **29**(13): p. i326-34.
985. Compeau, P.E., Pevzner, P.A., and Tesler, G., *How to apply de Bruijn graphs to genome assembly*. Nat Biotechnol, 2011. **29**(11): p. 987-91.
986. Altmann, A., et al., *A beginners guide to SNP calling from high-throughput DNA-sequencing data*. Hum Genet, 2012. **131**(10): p. 1541-54.
987. Albers, C.A., et al., *Dindel: accurate indel calls from short-read data*. Genome Res, 2011. **21**(6): p. 961-73.
988. Luke, C. and Frenzel, P., *Potential of pmoA amplicon pyrosequencing for methanotroph diversity studies*. Appl Environ Microbiol, 2011. **77**(17): p. 6305-9.

989. Neuman, J.A., Isakov, O., and Shomron, N., *Analysis of insertion-deletion from deep-sequencing data: software evaluation for optimal detection*. Brief Bioinform, 2013. **14**(1): p. 46-55.
990. Krawitz, P., et al., *Microindel detection in short-read sequence data*. Bioinformatics, 2010. **26**(6): p. 722-9.
991. Ossowski, S., et al., *Sequencing of natural strains of Arabidopsis thaliana with short reads*. Genome Res, 2008. **18**(12): p. 2024-33.
992. Abecasis, G.R., et al., *A map of human genome variation from population-scale sequencing*. Nature, 2010. **467**(7319): p. 1061-73.
993. Korbel, J.O., et al., *Paired-end mapping reveals extensive structural variation in the human genome*. Science, 2007. **318**(5849): p. 420-6.
994. Alkan, C., Coe, B.P., and Eichler, E.E., *Genome structural variation discovery and genotyping*. Nat Rev Genet, 2011. **12**(5): p. 363-76.
995. Le Scouarnec, S. and Gribble, S.M., *Characterising chromosome rearrangements: recent technical advances in molecular cytogenetics*. Heredity (Edinb), 2012. **108**(1): p. 75-85.
996. Karakoc, E., et al., *Detection of structural variants and indels within exome data*. Nat Methods, 2012. **9**(2): p. 176-8.
997. Krumm, N., et al., *Copy number variation detection and genotyping from exome sequence data*. Genome Res, 2012. **22**(8): p. 1525-32.
998. Menelaou, A. and Marchini, J., *Genotype calling and phasing using next-generation sequencing reads and a haplotype scaffold*. Bioinformatics, 2013. **29**(1): p. 84-91.
999. Duan, J., et al., *Comparative studies of copy number variation detection methods for next-generation sequencing technologies*. PLoS One, 2013. **8**(3): p. e59128.
1000. Bellos, E., Johnson, M.R., and M. Coin, M.J., *cnvHiTSeq: integrative models for high-resolution copy number variation detection and genotyping using population sequencing data*. Genome Biol, 2012. **13**(12): p. R120.
1001. Hedges, D.J., et al., *Comparison of three targeted enrichment strategies on the SOLiD sequencing platform*. PLoS One, 2011. **6**(4): p. e18595.
1002. Benjamini, Y. and Speed, T.P., *Summarizing and correcting the GC content bias in high-throughput sequencing*. Nucleic Acids Res, 2012. **40**(10): p. e72.
1003. Guo, Y., et al., *Exome sequencing generates high quality data in non-target regions*. BMC Genomics, 2012. **13**: p. 194.
1004. Kim, S.Y., et al., *Design of association studies with pooled or un-pooled next-generation sequencing data*. Genet Epidemiol, 2010. **34**(5): p. 479-91.
1005. O'Rawe, J., et al., *Low concordance of multiple variant-calling pipelines: practical implications for exome and genome sequencing*. Genome Med, 2013. **5**(3): p. 28.
1006. Shigemizu, D., et al., *A practical method to detect SNVs and indels from whole genome and exome sequencing data*. Sci Rep, 2013. **3**: p. 2161.
1007. Yu, X. and S. Sun, *Comparing a few SNP calling algorithms using low-coverage sequencing data*. BMC Bioinformatics, 2013. **14**(1): p. 274.
1008. Harismendy, O., et al., *Evaluation of next generation sequencing platforms for population targeted sequencing studies*. Genome Biol, 2009. **10**(3): p. R32.
1009. SAMtools. *Manual Reference Pages - samtools (1)*. samtools.sourceforge.net/samtools. Accessed online 26/09/13.
1010. The SAM/BAM Format Specification Working Group, *Sequence Alignment/Map Format Specification*, 2013: <http://samtools.sourceforge.net/SAMv1.pdf>. Accessed online 26/09/13.
1011. Jia, P., et al., *Consensus rules in variant detection from next-generation sequencing data*. PLoS One, 2012. **7**(6): p. e38470.

1012. Bainbridge, M.N., et al., *Targeted enrichment beyond the consensus coding DNA sequence exome reveals exons with higher variant densities*. *Genome Biol*, 2011. **12**(7): p. R68.
1013. Freudenberg-Hua, Y., et al., *Single nucleotide variation analysis in 65 candidate genes for CNS disorders in a representative sample of the European population*. *Genome Res*, 2003. **13**(10): p. 2271-6.
1014. Yang, Z. and Nielsen, R., *Synonymous and nonsynonymous rate variation in nuclear genes of mammals*. *J Mol Evol*, 1998. **46**(4): p. 409-18.
1015. Meacham, F., et al., *Identification and correction of systematic error in high-throughput sequence data*. *BMC Bioinformatics*, 2011. **12**: p. 451.
1016. Bell, C.J., et al., *Carrier testing for severe childhood recessive diseases by next-generation sequencing*. *Sci Transl Med*, 2011. **3**(65): p. 65ra4.
1017. Thomas, P.D. and Kejariwal, A., *Coding single-nucleotide polymorphisms associated with complex vs. Mendelian disease: evolutionary evidence for differences in molecular effects*. *Proc Natl Acad Sci USA*, 2004. **101**(43): p. 15398-403.
1018. Glazier, A.M., Nadeau, J.H., and Aitman, T.J., *Finding genes that underlie complex traits*. *Science*, 2002. **298**(5602): p. 2345-9.
1019. Bansal, V., et al., *Statistical analysis strategies for association studies involving rare variants*. *Nat Rev Genet*, 2010. **11**(11): p. 773-85.
1020. Abifadel, M., et al., *Mutations and polymorphisms in the proprotein convertase subtilisin kexin 9 (PCSK9) gene in cholesterol metabolism and disease*. *Hum Mutat*, 2009. **30**(4): p. 520-9.
1021. Benn, M., *Apolipoprotein B levels, APOB alleles, and risk of ischemic cardiovascular disease in the general population, a review*. *Atherosclerosis*, 2009. **206**(1): p. 17-30.
1022. Price, A.L., et al., *Pooled association tests for rare variants in exon-resequencing studies*. *Am J Hum Genet*, 2010. **86**(6): p. 832-8.
1023. Madsen, B.E. and Browning, S.R., *A groupwise association test for rare mutations using a weighted sum statistic*. *PLoS Genet*, 2009. **5**(2): p. e1000384.
1024. Hoh, J., Wille, A., and Ott, J., *Trimming, weighting, and grouping SNPs in human case-control association studies*. *Genome Res*, 2001. **11**(12): p. 2115-9.
1025. Bhatia, G., et al., *A covering method for detecting genetic associations between rare variants and common phenotypes*. *PLoS Comput Biol*, 2010. **6**(10): p. e1000954.
1026. Harismendy, O., et al., *Population sequencing of two endocannabinoid metabolic genes identifies rare and common regulatory variants associated with extreme obesity and metabolite level*. *Genome Biol*, 2010. **11**(11): p. R118.
1027. Hoischen, A., et al., *De novo mutations of SETBP1 cause Schinzel-Giedion syndrome*. *Nat Genet*, 2010. **42**(6): p. 483-5.
1028. Simpson, M.A., et al., *Mutations in NOTCH2 cause Hajdu-Cheney syndrome, a disorder of severe and progressive bone loss*. *Nat Genet*, 2011. **43**(4): p. 303-5.
1029. Ng, S.B., et al., *Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome*. *Nat Genet*, 2010. **42**(9): p. 790-3.
1030. O'Roak, B.J., et al., *Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations*. *Nature*, 2012. **485**(7397): p. 246-50.
1031. Sanders, S.J., et al., *De novo mutations revealed by whole-exome sequencing are strongly associated with autism*. *Nature*, 2012. **485**(7397): p. 237-41.
1032. Gratten, J., et al., *Interpreting the role of de novo protein-coding mutations in neuropsychiatric disease*. *Nat Genet*, 2013. **45**(3): p. 234-8.
1033. Stitzel, N.O., Kiezun, A., and Sunyaev, S., *Computational and statistical approaches to analyzing variants identified by exome sequencing*. *Genome Biol*, 2011. **12**(9): p. 227.

1034. Kanehisa, M., et al., *KEGG for integration and interpretation of large-scale molecular data sets*. Nucleic Acids Res, 2012. **40**(Database issue): p. D109-14.
1035. Keinan, A., et al., *Measurement of the human allele frequency spectrum demonstrates greater genetic drift in East Asians than in Europeans*. Nat Genet, 2007. **39**(10): p. 1251-5.
1036. Liu, Q., Nicolae, D.L., and Chen, L.S., *Marbled inflation from population structure in gene-based association studies with rare variants*. Genet Epidemiol, 2013. **37**(3): p. 286-92.
1037. He, H., et al., *Effect of population stratification analysis on false-positive rates for common and rare variants*. BMC Proc, 2011. **5 Suppl 9**: p. S116.
1038. O'Connor, T.D., et al., *Fine-scale patterns of population stratification confound rare variant association tests*. PLoS One, 2013. **8**(7): p. e65834.
1039. Tintle, N., et al., *Inflated type I error rates when using aggregation methods to analyze rare variants in the 1000 Genomes Project exon sequencing data in unrelated individuals: summary results from Group 7 at Genetic Analysis Workshop 17*. Genet Epidemiol, 2011. **35 Suppl 1**: p. S56-60.
1040. Ionita-Laza, I., et al., *Family-based association tests for sequence data, and comparisons with population-based association tests*. Eur J Hum Genet, 2013. **21**(10): p. 1158-62.
1041. Mathieson, I. and McVean, G., *Differential confounding of rare and common variants in spatially structured populations*. Nat Genet, 2012. **44**(3): p. 243-6.
1042. Heinrich, V., et al., *Estimating exome genotyping accuracy by comparing to data from large scale sequencing projects*. Genome Med, 2013. **5**(7): p. 69.
1043. Bainbridge, M.N., et al., *Whole exome capture in solution with 3 Gbp of data*. Genome Biol, 2010. **11**(6): p. R62.
1044. Knudsen, B. and Miyamoto, M.M., *Accurate and fast methods to estimate the population mutation rate from error prone sequences*. BMC Bioinformatics, 2009. **10**: p. 247.
1045. Parla, J.S., et al., *A comparative analysis of exome capture*. Genome Biol, 2011. **12**(9): p. R97.
1046. Indap, A.R., et al., *Variant discovery in targeted resequencing using whole genome amplified DNA*. BMC Genomics, 2013. **14**: p. 468.
1047. Fay, J.C., Wyckoff, G.J., and Wu, C.I., *Positive and negative selection on the human genome*. Genetics, 2001. **158**(3): p. 1227-34.
1048. Nejentsev, S., et al., *Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes*. Science, 2009. **324**(5925): p. 387-9.
1049. Fearnhead, N.S., et al., *Multiple rare variants in different genes account for multifactorial inherited susceptibility to colorectal adenomas*. Proc Natl Acad Sci U S A, 2004. **101**(45): p. 15992-7.
1050. Wu, M.C., et al., *Rare-variant association testing for sequencing data with the sequence kernel association test*. Am J Hum Genet, 2011. **89**(1): p. 82-93.
1051. Perez, C.A., et al., *A transient receptor potential channel expressed in taste receptor cells*. Nat Neurosci, 2002. **5**(11): p. 1169-76.
1052. Brixel, L.R., et al., *TRPM5 regulates glucose-stimulated insulin secretion*. Pflugers Arch, 2010. **460**(1): p. 69-76.
1053. Massa, M.L., Gagliardino, J.J., and Francini, F., *Liver glucokinase: An overview on the regulatory mechanisms of its activity*. IUBMB Life, 2011. **63**(1): p. 1-6.
1054. Shiota, C., et al., *Nuclear import of hepatic glucokinase depends upon glucokinase regulatory protein, whereas export is due to a nuclear export signal sequence in glucokinase*. J Biol Chem, 1999. **274**(52): p. 37125-30.

1055. Turner, J.G., Dawson, J., and Sullivan, D.M., *Nuclear export of proteins and drug resistance in cancer*. Biochem Pharmacol, 2012. **83**(8): p. 1021-32.
1056. Turner, J.G. and Sullivan, D.M., *CRM1-mediated nuclear export of proteins and drug resistance in cancer*. Curr Med Chem, 2008. **15**(26): p. 2648-55.
1057. Wright, K.J., et al., *An ARL3-UNC119-RP2 GTPase cycle targets myristoylated NPHP3 to the primary cilium*. Genes Dev, 2011. **25**(22): p. 2347-60.
1058. Gardin, A. and White, J., *The Sanger Mouse Genetics Programme: High Throughput Characterisation of Knockout Mice*. Acta Ophthalmologica, 2011. **89**(s248): p. 0.
1059. Kong, C., et al., *Ubiquitination and degradation of the hominoid-specific oncoprotein TBC1D3 is mediated by CUL7 E3 ligase*. PLoS One, 2012. **7**(9): p. e46485.
1060. Wainszelbaum, M.J., et al., *TBC1D3, a hominoid-specific gene, delays IRS-1 degradation and promotes insulin signaling by modulating p70 S6 kinase activity*. PLoS One, 2012. **7**(2): p. e31225.
1061. Levinson, D.F., et al., *Copy number variants in schizophrenia: confirmation of five previous findings and new evidence for 3q29 microdeletions and VIPR2 duplications*. Am J Psychiatry, 2011. **168**(3): p. 302-16.
1062. Lam, H.Y., et al., *Performance comparison of whole-genome sequencing platforms*. Nat Biotechnol, 2012. **30**(1): p. 78-82.
1063. Burdon, K.P., et al., *Genome-wide association study identifies susceptibility loci for open angle glaucoma at TMCO1 and CDKN2B-AS1*. Nat Genet, 2011. **43**(6): p. 574-8.
1064. Gondro, C., et al., *Quality control for genome-wide association studies*. Methods in Molecular Biology, 2013. **1019**: p. 129-147.
1065. Price, A.L., et al., *Principal components analysis corrects for stratification in genome-wide association studies*. Nat Genet, 2006. **38**(8): p. 904-9.
1066. Leek, J.T., et al., *Tackling the widespread and critical impact of batch effects in high-throughput data*. Nat Rev Genet, 2010. **11**(10): p. 733-9.
1067. Kumar, R., et al., *Inhibitors of renal epithelial phosphate transport in tumor-induced osteomalacia and uremia*. Proc Assoc Am Physicians, 1995. **107**(3): p. 296-305.
1068. Mosavi, L.K., et al., *The ankyrin repeat as molecular architecture for protein recognition*. Protein Sci, 2004. **13**(6): p. 1435-48.
1069. Cross, T.G., et al., *Serine/threonine protein kinases and apoptosis*. Exp Cell Res, 2000. **256**(1): p. 34-41.
1070. Sparks, L.M., et al., *Remodeling lipid metabolism and improving insulin responsiveness in human primary myotubes*. PLoS One, 2011. **6**(7): p. e21068.
1071. Lee, J.C. and Peter, M.E., *Regulation of apoptosis by ubiquitination*. Immunol Rev, 2003. **193**: p. 39-47.
1072. Steenweg, M.E., et al., *An overview of L-2-Hydroxyglutarate dehydrogenase gene (L2HGDH) variants: a genotype-phenotype study*. Hum Mutat, 2010. **31**: p. 380-390.
1073. Bogerd, H.P., et al., *Inhibition of human immunodeficiency virus Rev and human T-cell leukemia virus Rex function, but not Mason-Pfizer monkey virus constitutive transport element activity, by a mutant human nucleoporin targeted to Crm1*. J Virol, 1998. **72**(11): p. 8627-35.
1074. Hayakawa, K., Arai, K., and Lo, E.H., *Role of ERK map kinase and CRM1 in IL-1beta-stimulated release of HMGB1 from cortical astrocytes*. Glia, 2010. **58**(8): p. 1007-15.
1075. Seternes, O.M., et al., *Both binding and activation of p38 mitogen-activated protein kinase (MAPK) play essential roles in regulation of the nucleocytoplasmic distribution of MAPK-activated protein kinase 5 by cellular stress*. Mol Cell Biol, 2002. **22**(20): p. 6931-45.

1076. Kudo, N., et al., *A novel nuclear export signal sensitive to oxidative stress in the fission yeast transcription factor Pap1*. J Biol Chem, 1999. **274**(21): p. 15151-8.
1077. Walker, C.J., et al., *Preclinical and clinical efficacy of XPO1/CRM1 inhibition by the karyopherin inhibitor KPT-330 in Ph+ leukemias*. Blood, 2013. **122**(17): p. 3034-44.
1078. Smith, J.J., et al., *Apoptotic regulation by the Crk adapter protein mediated by interactions with Wee1 and Crm1/exportin*. Mol Cell Biol, 2002. **22**(5): p. 1412-23.
1079. Mukhtar, M., Stubbs, M., and Agius, L., *Evidence for glucose and sorbitol-induced nuclear export of glucokinase regulatory protein in hepatocytes*. FEBS Lett, 1999. **462**(3): p. 453-8.
1080. Millar, C.A., et al., *Evidence for a role for ADP-ribosylation factor 6 in insulin-stimulated glucose transporter-4 (GLUT4) trafficking in 3T3-L1 adipocytes*. J Biol Chem, 1999. **274**(25): p. 17619-25.
1081. Colsoul, B., et al., *Loss of high-frequency glucose-induced Ca²⁺ oscillations in pancreatic islets correlates with impaired glucose tolerance in Trpm5^{-/-} mice*. Proc Natl Acad Sci U S A, 2010. **107**(11): p. 5208-13.
1082. Ashcroft, F.M., Harrison, D.E., and Ashcroft, S.J., *Glucose induces closure of single potassium channels in isolated rat pancreatic beta-cells*. Nature, 1984. **312**(5993): p. 446-8.
1083. Zhang, Y., et al., *Coding of sweet, bitter, and umami tastes: different receptor cells sharing similar signaling pathways*. Cell, 2003. **112**(3): p. 293-301.
1084. Ren, X., et al., *Nutrient selection in the absence of taste receptor signaling*. J Neurosci, 2010. **30**(23): p. 8012-23.
1085. Satoh, T., *The Ras Superfamily of Small GTP-Binding Proteins in Glucose Transporter Type 4-Mediated Glucose Uptake in Insulin-Responsive Tissues*, in *Glucose Uptake*, C.C. Johnson and D.B. Williams, Editors. 2013, Nova Science Publishers, Inc. p. 235-252.
1086. Vallar, L., Biden, T.J., and Wollheim, C.B., *Guanine nucleotides induce Ca²⁺-independent insulin secretion from permeabilized RINm5F cells*. J Biol Chem, 1987. **262**(11): p. 5049-56.
1087. Kibbey, R.G., et al., *Mitochondrial GTP regulates glucose-stimulated insulin secretion*. Cell Metab, 2007. **5**(4): p. 253-64.
1088. Meshkini, A., Yazdanparast, R., and Nouri, K., *Intracellular GTP level determines cell's fate toward differentiation and apoptosis*. Toxicol Appl Pharmacol, 2011. **253**(3): p. 188-96.
1089. Coleman, M.L. and Olson, M.F., *Rho GTPase signalling pathways in the morphological changes associated with apoptosis*. Cell Death Differ, 2002. **9**(5): p. 493-504.
1090. Huo, J., et al., *Activation of caspase-2 mediates the apoptosis induced by GTP-depletion in insulin-secreting (HIT-T15) cells*. Endocrinology, 2002. **143**(5): p. 1695-704.
1091. Babatz, T.D., et al., *Copy number and sequence variants implicate APBA2 as an autism candidate gene*. Autism Res, 2009. **2**(6): p. 359-64.
1092. Borg, J.P., et al., *The X11alpha protein slows cellular amyloid precursor protein processing and reduces Abeta40 and Abeta42 secretion*. J Biol Chem, 1998. **273**(24): p. 14761-6.
1093. Niu, T.K., et al., *Dynamics of GBF1, a Brefeldin A-sensitive Arf1 exchange factor at the Golgi*. Mol Biol Cell, 2005. **16**(3): p. 1213-22.
1094. Miyamoto, T., et al., *AMP-activated protein kinase phosphorylates Golgi-specific brefeldin A resistance factor 1 at Thr1337 to induce disassembly of Golgi apparatus*. J Biol Chem, 2008. **283**(7): p. 4430-8.

1095. Yang, J., et al., *Control of cyclin B1 localization through regulated binding of the nuclear export factor CRM1*. Genes Dev, 1998. **12**(14): p. 2131-43.
1096. Gulacsi, A. and Lillien, L., *Sonic hedgehog and bone morphogenetic protein regulate interneuron development from dorsal telencephalic progenitors in vitro*. J Neurosci, 2003. **23**(30): p. 9862-72.
1097. Thibert, C., et al., *Inhibition of neuroepithelial patched-induced apoptosis by sonic hedgehog*. Science, 2003. **301**(5634): p. 843-6.
1098. Belloni, E., et al., *Identification of Sonic hedgehog as a candidate gene responsible for holoprosencephaly*. Nat Genet, 1996. **14**(3): p. 353-6.
1099. Roessler, E., et al., *Mutations in the human Sonic Hedgehog gene cause holoprosencephaly*. Nat Genet, 1996. **14**(3): p. 357-60.
1100. Ribeiro, L.A., Murray, J.C., and Richieri-Costa, A., *PTCH mutations in four Brazilian patients with holoprosencephaly and in one with holoprosencephaly-like features and normal MRI*. Am J Med Genet A, 2006. **140**(23): p. 2584-6.
1101. Rahimov, F., et al., *GLI2 mutations in four Brazilian patients: how wide is the phenotypic spectrum?* Am J Med Genet A, 2006. **140**(23): p. 2571-6.
1102. Derwinska, K., et al., *PTCH1 duplication in a family with microcephaly and mild developmental delay*. Eur J Hum Genet, 2009. **17**(2): p. 267-71.
1103. Hahn, H., et al., *Mutations of the human homolog of Drosophila patched in the nevoid basal cell carcinoma syndrome*. Cell, 1996. **85**(6): p. 841-51.
1104. Johnson, R.L., et al., *Human homolog of patched, a candidate gene for the basal cell nevus syndrome*. Science, 1996. **272**(5268): p. 1668-71.
1105. Wolf, I., et al., *Unmasking of epigenetically silenced genes reveals DNA promoter methylation and reduced expression of PTCH in breast cancer*. Breast Cancer Res Treat, 2007. **105**(2): p. 139-55.
1106. Pritchard, J.I. and Olson, J.M., *Methylation of PTCH1, the Patched-1 gene, in a panel of primary medulloblastomas*. Cancer Genet Cytogenet, 2008. **180**(1): p. 47-50.
1107. Saldanha, G., *The Hedgehog signalling pathway and cancer*. J Pathol, 2001. **193**(4): p. 427-32.
1108. Grachtchouk, M., et al., *Basal cell carcinomas in mice overexpressing Gli2 in skin*. Nat Genet, 2000. **24**(3): p. 216-7.
1109. Dahmane, N., et al., *Activation of the transcription factor Gli1 and the Sonic hedgehog signalling pathway in skin tumours*. Nature, 1997. **389**(6653): p. 876-81.
1110. Fan, H., et al., *Induction of basal cell carcinoma features in transgenic human skin expressing Sonic Hedgehog*. Nat Med, 1997. **3**(7): p. 788-92.
1111. Ming, J.E., et al., *Mutations in PATCHED-1, the receptor for SONIC HEDGEHOG, are associated with holoprosencephaly*. Hum Genet, 2002. **110**(4): p. 297-301.
1112. Sims, J.R., et al., *Sonic hedgehog regulates ischemia/hypoxia-induced neural progenitor proliferation*. Stroke, 2009. **40**(11): p. 3618-26.
1113. Dokucu, A.I., et al., *The effects of molsidomine on hypoxia inducible factor alpha and Sonic hedgehog in testicular ischemia/reperfusion injury in rats*. Int Urol Nephrol, 2009. **41**(1): p. 101-8.
1114. Bijlsma, M.F., et al., *Hypoxia induces a hedgehog response mediated by HIF-1alpha*. J Cell Mol Med, 2009. **13**(8B): p. 2053-60.
1115. Ozturk, H., Tuncer, M.C., and Buyukbayram, H., *Nitric oxide regulates expression of sonic hedgehog and hypoxia-inducible factor-1alpha in an experimental model of kidney ischemia-reperfusion*. Ren Fail, 2007. **29**(3): p. 249-56.
1116. Hwang, J.M., et al., *Hypoxia-induced compensatory effect as related to Shh and HIF-1alpha in ischemia embryo rat heart*. Mol Cell Biochem, 2008. **311**(1-2): p. 179-87.
1117. Dai, R.L., et al., *Sonic hedgehog protects cortical neurons against oxidative stress*. Neurochem Res, 2011. **36**(1): p. 67-75.

1118. Dai, R., et al., *Involvement of PI3K/Akt pathway in the neuroprotective effect of Sonic hedgehog on cortical neurons under oxidative stress*. J Huazhong Univ Sci Technolog Med Sci, 2012. **32**(6): p. 856-60.
1119. Wu, J., Zhang, S., and Sun, X., *Neuroprotective effect of upregulated sonic Hedgehog in retinal ganglion cells following chronic ocular hypertension*. Invest Ophthalmol Vis Sci, 2010. **51**(6): p. 2986-92.
1120. Miao, N., et al., *Sonic hedgehog promotes the survival of specific CNS neuron populations and protects these cells from toxic insult In vitro*. J Neurosci, 1997. **17**(15): p. 5891-9.
1121. Wu, C.L., et al., *Sonic hedgehog mediates BDNF-induced neuroprotection against mitochondrial inhibitor 3-nitropropionic acid*. Biochem Biophys Res Commun, 2009. **385**(1): p. 112-7.
1122. Al-Ayadhi, L.Y., *Relationship between Sonic hedgehog protein, brain-derived neurotrophic factor and oxidative stress in autism spectrum disorders*. Neurochem Res, 2012. **37**(2): p. 394-400.
1123. Petti, A.A., et al., *Survival of starving yeast is correlated with oxidative stress response and nonrespiratory mitochondrial function*. Proc Natl Acad Sci USA, 2011. **108**(45): p. E1089-98.
1124. Henriksen, L., et al., *Internalization mechanisms of the epidermal growth factor receptor after activation with different ligands*. PLoS One, 2013. **8**(3): p. e58148.
1125. Roepstorff, K., et al., *Differential effects of EGFR ligands on endocytic sorting of the receptor*. Traffic, 2009. **10**(8): p. 1115-27.
1126. Oda, K., et al., *A comprehensive pathway map of epidermal growth factor receptor signaling*. Mol Syst Biol, 2005. **1**: p. 2005.0010.
1127. Yarden, Y. and Sliwkowski, M.X., *Untangling the ErbB signalling network*. Nat Rev Mol Cell Biol, 2001. **2**(2): p. 127-37.
1128. Sibilia, M., et al., *A strain-independent postnatal neurodegeneration in mice lacking the EGF receptor*. EMBO J, 1998. **17**(3): p. 719-31.
1129. Wikstrand, C.J., et al., *The class III variant of the epidermal growth factor receptor (EGFRvIII): characterization and utilization as an immunotherapeutic target*. J Neurovirol, 1998. **4**(2): p. 148-58.
1130. Opanashuk, L.A., et al., *Heparin-binding epidermal growth factor-like growth factor in hippocampus: modulation of expression by seizures and anti-excitotoxic action*. J Neurosci, 1999. **19**(1): p. 133-46.
1131. Fallon, J., et al., *In vivo induction of massive proliferation, directed migration, and differentiation of neural cells in the adult mammalian brain*. Proc Natl Acad Sci U S A, 2000. **97**(26): p. 14686-91.
1132. Koprivica, V., et al., *EGFR activation mediates inhibition of axon regeneration by myelin and chondroitin sulfate proteoglycans*. Science, 2005. **310**(5745): p. 106-10.
1133. Hendriks, B.S., et al., *Decreased internalisation of erbB1 mutants in lung cancer is linked with a mechanism conferring sensitivity to gefitinib*. Syst Biol (Stevenage), 2006. **153**(6): p. 457-66.
1134. Rogers, C., et al., *EGF receptor (ERBB1) abundance in adipose tissue is reduced in insulin-resistant and type 2 diabetic women*. J Clin Endocrinol Metab, 2012. **97**(3): p. E329-40.
1135. Prada, P.O., et al., *EGFR tyrosine kinase inhibitor (PD153035) improves glucose tolerance and insulin action in high-fat diet-fed mice*. Diabetes, 2009. **58**(12): p. 2910-9.
1136. Lehner, R. and Kuksis, A., *Biosynthesis of triacylglycerols*. Prog Lipid Res, 1996. **35**(2): p. 169-201.

1137. Nye, C., et al., *Reassessing triglyceride synthesis in adipose tissue*. Trends Endocrinol Metab, 2008. **19**(10): p. 356-61.
1138. Homko, C.J., Cheung, P., and Boden, G., *Effects of free fatty acids on glucose uptake and utilization in healthy women*. Diabetes, 2003. **52**(2): p. 487-91.
1139. Saini, K.S., et al., *Targeting the PI3K/AKT/mTOR and Raf/MEK/ERK pathways in the treatment of breast cancer*. Cancer Treat Rev, 2013. **39**(8): p. 935-46.
1140. Tao, Y., et al., *Mechanisms of disease: signaling of the insulin-like growth factor 1 receptor pathway--therapeutic perspectives in cancer*. Nat Clin Pract Oncol, 2007. **4**(10): p. 591-602.
1141. Zhao, Y., et al., *Interactions between SIRT1 and MAPK/ERK regulate neuronal apoptosis induced by traumatic brain injury in vitro and in vivo*. Exp Neurol, 2012. **237**(2): p. 489-98.
1142. Wang, Q., et al., *Protein kinase B/Akt participates in GLUT4 translocation by insulin in L6 myoblasts*. Mol Cell Biol, 1999. **19**(6): p. 4008-18.
1143. Deberardinis, R.J., Lum, J.J., and Thompson, C.B., *Phosphatidylinositol 3-kinase-dependent modulation of carnitine palmitoyltransferase 1A expression regulates lipid metabolism during hematopoietic cell growth*. J Biol Chem, 2006. **281**(49): p. 37372-80.
1144. Klement, R.J. and Kammerer, U., *Is there a role for carbohydrate restriction in the treatment and prevention of cancer?* Nutr Metab (Lond), 2011. **8**: p. 75.
1145. Laplante, M. and Sabatini, D.M., *mTOR signaling at a glance*. J Cell Sci, 2009. **122**(Pt 20): p. 3589-94.
1146. Ouban, A., et al., *Expression and distribution of insulin-like growth factor-1 receptor in human carcinomas*. Hum Pathol, 2003. **34**(8): p. 803-8.
1147. Khandwala, H.M., et al., *The effects of insulin-like growth factors on tumorigenesis and neoplastic growth*. Endocr Rev, 2000. **21**(3): p. 215-44.
1148. Tsuchiya, N., et al., *CA repeat polymorphism in the insulin-like growth factor-I gene is associated with increased risk of prostate cancer and benign prostatic hyperplasia*. Int J Oncol, 2005. **26**(1): p. 225-31.
1149. Tsuchiya, N., et al., *Impact of IGF-I and CYP19 gene polymorphisms on the survival of patients with metastatic prostate cancer*. J Clin Oncol, 2006. **24**(13): p. 1982-9.
1150. Cheng, I., et al., *Common genetic variation in IGF1 and prostate cancer risk in the Multiethnic Cohort*. J Natl Cancer Inst, 2006. **98**(2): p. 123-34.
1151. Pantaleo, M.A., et al., *The emerging role of insulin-like growth factor 1 receptor (IGF1r) in gastrointestinal stromal tumors (GISTs)*. J Transl Med, 2010. **8**: p. 117.
1152. Albanell, J. and Baselga, J., *Unraveling resistance to trastuzumab (Herceptin): insulin-like growth factor-I receptor, a new suspect*. J Natl Cancer Inst, 2001. **93**(24): p. 1830-2.
1153. Lu, Y., et al., *Insulin-like growth factor-I receptor signaling and resistance to trastuzumab (Herceptin)*. J Natl Cancer Inst, 2001. **93**(24): p. 1852-7.
1154. Jones, H.E., et al., *Insulin-like growth factor-I receptor signalling and acquired resistance to gefitinib (ZD1839; Iressa) in human breast and prostate cancer cells*. Endocr Relat Cancer, 2004. **11**(4): p. 793-814.
1155. Chakravarti, A., Loeffler, J.S., and Dyson, N.J., *Insulin-like growth factor receptor I mediates resistance to anti-epidermal growth factor receptor therapy in primary human glioblastoma cells through continued activation of phosphoinositide 3-kinase signaling*. Cancer Res, 2002. **62**(1): p. 200-7.
1156. Seyfried, T.N., et al., *Role of glucose and ketone bodies in the metabolic control of experimental brain cancer*. Br J Cancer, 2003. **89**(7): p. 1375-82.

1157. Aft, R.L., F.W. Zhang, and D. Gius, *Evaluation of 2-deoxy-D-glucose as a chemotherapeutic agent: mechanism of cell death*. Br J Cancer, 2002. **87**(7): p. 805-12.
1158. Mukherjee, P., et al., *Dietary restriction reduces angiogenesis and growth in an orthotopic mouse brain tumour model*. Br J Cancer, 2002. **86**(10): p. 1615-21.
1159. Mulrooney, T.J., et al., *Influence of caloric restriction on constitutive expression of NF-kappaB in an experimental mouse astrocytoma*. PLoS One, 2011. **6**(3): p. e18085.
1160. van Heemst, D., *Insulin, IGF-1 and longevity*. Aging Dis, 2010. **1**(2): p. 147-57.
1161. Anton, S. and Leeuwenburgh, C., *Fasting or caloric restriction for healthy aging*. Exp Gerontol, 2013. **48**(10): p. 1003-5.
1162. Rizza, W., Veronese, N., and Fontana, L., *What are the roles of calorie restriction and diet quality in promoting healthy longevity?* Ageing Res Rev, 2013. **13C**: p. 38-45.
1163. Kinzig, K.P., Honors, M.A., and Hargrave, S.L., *Insulin sensitivity and glucose tolerance are altered by maintenance on a ketogenic diet*. Endocrinology, 2010. **151**: p. 3105-14.
1164. Samaha, F.F., et al., *A low-carbohydrate as compared with a low-fat diet in severe obesity*. N Engl J Med, 2003. **348**(21): p. 2074-81.
1165. Dushay, J., et al., *Increased fibroblast growth factor 21 in obesity and nonalcoholic fatty liver disease*. Gastroenterology, 2010. **139**(2): p. 456-63.
1166. Johnston, B.M., et al., *Insulin-like growth factor-1 is a potent neuronal rescue agent after hypoxic-ischemic injury in fetal lambs*. J Clin Invest, 1996. **97**(2): p. 300-8.
1167. Jin, Z., et al., *Insulin reduces neuronal excitability by turning on GABA(A) channels that generate tonic current*. PLoS One, 2011. **6**(1): p. e16188.
1168. O'Malley, D. and Harvey, J., *Insulin activates native and recombinant large conductance Ca(2+)-activated potassium channels via a mitogen-activated protein kinase-dependent process*. Mol Pharmacol, 2004. **65**(6): p. 1352-63.
1169. O'Malley, D., Shanley, L.J., and Harvey, J., *Insulin inhibits rat hippocampal neurones via activation of ATP-sensitive K+ and large conductance Ca2+-activated K+ channels*. Neuropharmacology, 2003. **44**(7): p. 855-63.
1170. Llorens-Martin, M., Torres-Aleman, I., and Trejo, J.L., *Mechanisms mediating brain plasticity: IGF1 and adult hippocampal neurogenesis*. Neuroscientist, 2009. **15**(2): p. 134-48.
1171. Davila, D., et al., *Insulin and insulin-like growth factor I signalling in neurons*. Front Biosci, 2007. **12**: p. 3194-202.
1172. Casadesus, G., et al., *Modulation of hippocampal plasticity and cognitive behavior by short-term blueberry supplementation in aged rats*. Nutr Neurosci, 2004. **7**(5-6): p. 309-16.
1173. Houser, C.R., Huang, C.S., and Peng, Z., *Dynamic seizure-related changes in extracellular signal-regulated kinase activation in a mouse model of temporal lobe epilepsy*. Neuroscience, 2008. **156**(1): p. 222-37.
1174. Thomas, G.M. and Huganir, R.L., *MAPK cascade signalling and synaptic plasticity*. Nat Rev Neurosci, 2004. **5**(3): p. 173-83.
1175. Datta, S.R., et al., *Akt phosphorylation of BAD couples survival signals to the cell-intrinsic death machinery*. Cell, 1997. **91**(2): p. 231-41.
1176. Choi, Y.S., et al., *Status epilepticus-induced somatostatinergic hilar interneuron degeneration is regulated by striatal enriched protein tyrosine phosphatase*. J Neurosci, 2007. **27**(11): p. 2999-3009.
1177. Torres-Aleman, I., *Serum growth factors and neuroprotective surveillance: focus on IGF-1*. Mol Neurobiol, 2000. **21**(3): p. 153-60.

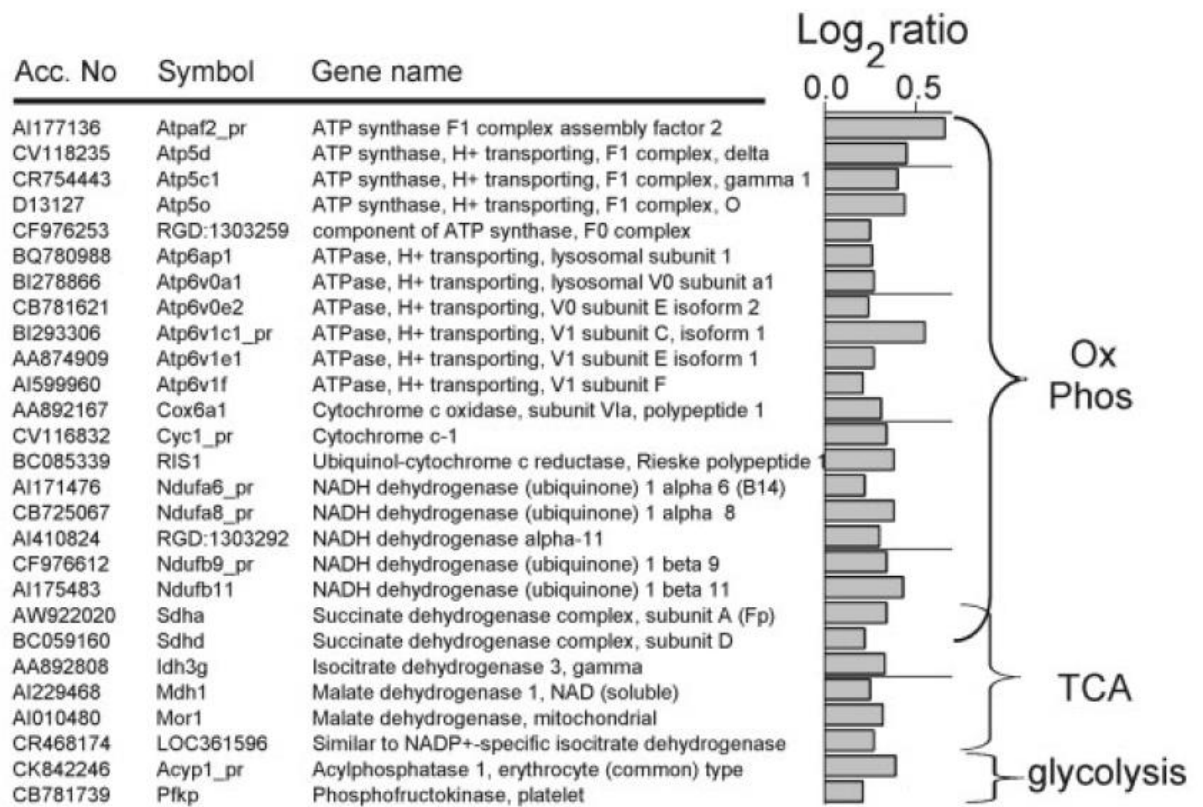
1178. Xing, C., et al., *Effects of insulin-like growth factor 1 on synaptic excitability in cultured rat hippocampal neurons*. *Exp Neurol*, 2007. **205**(1): p. 222-9.
1179. Blagosklonny, M.V., *Once again on rapamycin-induced insulin resistance and longevity: despite of or owing to*. *Aging (Albany NY)*, 2012. **4**(5): p. 350-8.
1180. Koffler, M. and Kisch, E.S., *Starvation diet and very-low-calorie diets may induce insulin resistance and overt diabetes mellitus*. *J Diabetes Complications*, 1996. **10**(2): p. 109-12.
1181. Kaklamani, V.G., et al., *Dietary fat and carbohydrates are independently associated with circulating insulin-like growth factor 1 and insulin-like growth factor-binding protein 3 concentrations in healthy adults*. *J Clin Oncol*, 1999. **17**(10): p. 3291-8.
1182. Dunn, S.E., et al., *Dietary restriction reduces insulin-like growth factor I levels, which modulates apoptosis, cell proliferation, and tumor progression in p53-deficient mice*. *Cancer Res*, 1997. **57**(21): p. 4667-72.
1183. Jiang, W., Zhu, Z., and Thompson, H.J., *Dietary energy restriction modulates the activity of AMP-activated protein kinase, Akt, and mammalian target of rapamycin in mammary carcinomas, mammary gland, and liver*. *Cancer Res*, 2008. **68**(13): p. 5492-9.
1184. Fontana, L., et al., *Long-term effects of calorie or protein restriction on serum IGF-1 and IGFBP-3 concentration in humans*. *Aging Cell*, 2008. **7**(5): p. 681-7.
1185. Raile, K., et al., *Glucose concentration and AMP-dependent kinase activation regulate expression of insulin receptor family members in rat islets and INS-1E beta cells*. *Diabetologia*, 2005. **48**(9): p. 1798-809.
1186. McDaniel, S.S., et al., *The ketogenic diet inhibits the mammalian target of rapamycin (mTOR) pathway*. *Epilepsia*, 2011. **52**(3): p. e7-11.
1187. Spulber, G., et al., *Growth dependence on insulin-like growth factor-1 during the ketogenic diet*. *Epilepsia*, 2009. **50**(2): p. 297-303.
1188. Sharma, N., et al., *Improved insulin sensitivity with calorie restriction does not require reduced JNK1/2, p38, or ERK1/2 phosphorylation in skeletal muscle of 9-month-old rats*. *Am J Physiol Regul Integr Comp Physiol*, 2012. **302**(1): p. R126-36.
1189. Dean, D.J., et al., *Calorie restriction increases cell surface GLUT-4 in insulin-stimulated skeletal muscle*. *Am J Physiol*, 1998. **275**(6 Pt 1): p. E957-64.
1190. Schenk, S., et al., *Sirt1 enhances skeletal muscle insulin sensitivity in mice during caloric restriction*. *J Clin Invest*, 2011. **121**(11): p. 4281-8.
1191. Puig, O. and Tjian, R., *Transcriptional feedback control of insulin receptor by dFOXO/FOXO1*. *Genes Dev*, 2005. **19**(20): p. 2435-46.
1192. Zbinden-Foncea, H., et al., *Contribution of nonesterified fatty acids to mitogen-activated protein kinase activation in human skeletal muscle during endurance exercise*. *Int J Sport Nutr Exerc Metab*, 2013. **23**(3): p. 201-9.
1193. Sharma, N., Castorena, C.M., and Cartee, G.D., *Tissue-specific responses of IGF-1/insulin and mTOR signaling in calorie restricted rats*. *PLoS One*, 2012. **7**(6): p. e38835.
1194. de la Monte, S.M. and Wands, J.R., *Review of insulin and insulin-like growth factor expression, signaling, and malfunction in the central nervous system: relevance to Alzheimer's disease*. *J Alzheimers Dis*, 2005. **7**(1): p. 45-61.
1195. Talbot, K., et al., *Demonstrated brain insulin resistance in Alzheimer's disease patients is associated with IGF-1 resistance, IRS-1 dysregulation, and cognitive decline*. *J Clin Invest*, 2012. **122**(4): p. 1316-38.
1196. Umegaki, H., *Insulin resistance in the brain: A new therapeutic target for Alzheimer's disease*. *J Diabetes Invest*, 2013. **4**: p. 150-151.
1197. Watson, G.S., et al., *Insulin increases CSF Abeta42 levels in normal older adults*. *Neurology*, 2003. **60**(12): p. 1899-903.

1198. Idris, I., Miller, D., and Page, S.R., *Familial hyperinsulinaemia associated with epilepsy and mental retardation--a syndrome of familial insulin resistance*. Diabet Med, 2004. **21**(6): p. 628-31.
1199. Atabek, M.E., et al., *Cohen syndrome with insulin resistance and seizure*. Pediatric Neurology, 2004. **30**(1): p. 61-3.
1200. Varlamis, S., et al., *Evaluation of Oral Glucose Tolerance Test in Children With Epilepsy*. J Child Neurol, 2013. **28**(11): p. 1437-1442.
1201. Ferre, P. and Fufelle, F., *Hepatic steatosis: a role for de novo lipogenesis and the transcription factor SREBP-1c*. Diabetes Obes Metab, 2010. **12 Suppl 2**: p. 83-92.
1202. Strable, M.S. and Ntambi, J.M., *Genetic control of de novo lipogenesis: role in diet-induced obesity*. Crit Rev Biochem Mol Biol, 2010. **45**(3): p. 199-214.
1203. Matsuzaka, T. and Shimano, H., *Elovl6: a new player in fatty acid metabolism and insulin sensitivity*. J Mol Med (Berl), 2009. **87**(4): p. 379-84.
1204. Li, L.O., et al., *Liver-specific loss of long chain acyl-CoA synthetase-1 decreases triacylglycerol synthesis and beta-oxidation and alters phospholipid fatty acid composition*. J Biol Chem, 2009. **284**(41): p. 27816-26.
1205. Freedman, D.S., et al., *Relation of body fat distribution to hyperinsulinemia in children and adolescents: the Bogalusa Heart Study*. Am J Clin Nutr, 1987. **46**(3): p. 403-10.
1206. Gower, B.A., et al., *Fat distribution and insulin response in prepubertal African American and white children*. Am J Clin Nutr, 1998. **67**(5): p. 821-7.
1207. Ku, C.Y., et al., *Racial differences in insulin secretion and sensitivity in prepubertal children: role of physical fitness and physical activity*. Obes Res, 2000. **8**(7): p. 506-15.
1208. Kodama, K., et al., *Ethnic differences in the relationship between insulin sensitivity and insulin response: a systematic review and meta-analysis*. Diabetes Care, 2013. **36**(6): p. 1789-96.
1209. Sullivan, P.F., Fan, C., and Perou, C.M., *Evaluating the comparability of gene expression in blood and brain*. Am J Med Genet B Neuropsychiatr Genet, 2006. **141B**(3): p. 261-8.
1210. Scassa, M.E., et al., *Hepatic nuclear factor 3 and nuclear factor 1 regulate 5-aminolevulinate synthase gene expression and are involved in insulin repression*. J Biol Chem, 2004. **279**(27): p. 28082-92.
1211. Taniguchi, C.M., Emanuelli, B., and Kahn, C.R., *Critical nodes in signalling pathways: insights into insulin action*. Nat Rev Mol Cell Biol, 2006. **7**(2): p. 85-96.

10 Appendices

Appendix 1.1

Fig 3B from Bough et al., 2006 [291].



Relative expression levels of 24 energy metabolism genes that were upregulated by the KD.

Each row corresponds to one gene. Genes associated with glycolysis, the tricarboxylic acid

(TCA) cycle, or oxidative phosphorylation (Ox Phos) are grouped.

Appendix 1.2

Taken from Supplemental Table 1, Bough et al., 2006 [291].

Access No.	Symbol	Log2 ratio (KD/Control)	Gene Name
CB717809	Ttc11_pr	0.39	Tetratricopeptide repeat domain 11
AI172214	Mrpl53_pr	0.19	Mitochondrial ribosomal protein L53
AI599839	LOC298684	0.66	Similar to Mrpl20 protein
AI170683	Oat	0.42	Ornithine aminotransferase
XM_215897	Acas2l_pr	0.39	Acetyl-Coenzyme A synthetase 2 (AMP forming)-like
AI177136	Atpaf2_pr	0.66	ATP synthase mitochondrial F1 complex assembly factor 2
CV118235	Atp5d	0.45	ATP synthase, H ⁺ transporting, mitochondrial F1 complex, delta subunit
CR754443	Atp5c1	0.4	ATP synthase, H ⁺ transporting, mitochondrial F1 complex, gamma 1
D13127	Atp5o	0.44	ATP synthase, H ⁺ transporting, mitochondrial F1 complex, O subunit
BI293306	Atp6v1c1_pr	0.55	ATPase, H ⁺ transporting, V1 subunit C, isoform 1
AA874909	Atp6v1e1	0.27	ATPase, H ⁺ transporting, V1 subunit E isoform 1
AI599960	Atp6v1f	0.21	ATPase, H ⁺ transporting, V1 subunit F
AA892167	Cox6a1	0.31	Cytochrome c oxidase, subunit VIa, polypeptide 1
CV116832	Cyc1_pr	0.34	Cytochrome c-1
AA892808	Idh3g	0.33	Isocitrate dehydrogenase 3, gamma
AI010480	Mor1	0.32	Malate dehydrogenase, mitochondrial
AI171476	Ndufa6_pr	0.22	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 6 (B14)
CB725067	Ndufa8_pr	0.38	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 8
CF976612	Ndufb9_pr	0.34	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 9
CF976253	RGD:1303259	0.25	Similar to CG6105-PA (component of ATP synthase, F0 complex)
CB327643	Slc25a4	0.23	Solute carrier family 25 (mitochondrial adenine nucleotide translocator 4)
CA340524	Slc25a3	0.35	Solute carrier family 25 (mitochondrial adenine nucleotide translocator 3)
AW922020	Sdha	0.34	Succinate dehydrogenase complex, subunit A, flavoprotein (Fp)
BC059160	Sdhb	0.22	Succinate dehydrogenase complex, subunit D, membrane protein
		0.26	similar to NADH-ubiquinone oxidoreductase 18 kDa subunit
BC085339	RIS1	0.38	Ubiquinol-cytochrome c reductase, Rieske iron-sulfur polypeptide 1
AI175483	Ndufb11	0.43	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 11

AA800237	Acadm	-0.39	Acetyl-Coenzyme A dehydrogenase, medium chain
BM390148	AcsI3	0.4	Acyl-CoA synthetase long-chain family member 3
AI176264	Dci	0.46	Dodecenoyl-coenzyme A delta isomerase
CK840087	Gpam	-0.41	Glycerol-3-phosphate acyltransferase, mitochondrial
AI102615	Hadhb	0.35	Hydroxyacyl-Coenzyme A dehydrogenase (trifunctional protein), beta
BF285297	—	0.31	Transcribed locus
NM_080689	Dnm1	0.28	Dynamin 1
AF019043	Dnm1l	0.29	Dynamin 1-like
BF282660	Timm9	0.36	Similar to Mitochondrial inner membrane translocase subunit TIM9 A
NM_152935	Tomm20	0.3	Translocase of outer mitochondrial membrane 20 homolog (yeast)
XM_215358	Mtch1_pr	0.2	Mitochondrial carrier homolog 1 (C. elegans)
CB325437	Immt_pr	0.43	Inner membrane protein, mitochondrial
AF095741	Mg87	1.04	Mg87 protein
BM383379	RGD:1303219	0.34	Similar to Protein CGI-51

Appendix 2.1



National Research Ethics Service

East Central London REC 1

South House, Block A
Royal Free Hospital
Pond Street
London
NW3 2QG

Telephone: 020 7794 0562

Professor Sanjay Sisodiya
UCL Institute of Neurology
National Hospital for Neurology and Neurosurgery
Queens Square
London
WC1N 3BG

08 March 2011

Dear Professor Sisodiya

Study Title: A Genetic Basis for Response to the Ketogenic Diet in Epilepsy
REC reference number: 11/H0721/4
Protocol number: 1.0

Thank you for your letter of 23 February 2011, responding to the Committee's request for further information on the above research and submitting revised documentation.

The further information has been considered on behalf of the Committee by the Vice-Chair.

Confirmation of ethical opinion

On behalf of the Committee, I am pleased to confirm a favourable ethical opinion for the above research on the basis described in the application form, protocol and supporting documentation as revised, subject to the conditions specified below.

Ethical review of research sites

The favourable opinion applies to all NHS sites taking part in the study, subject to management permission being obtained from the NHS/HSC R&D office prior to the start of the study (see 'Conditions of the favourable opinion' below).

The Committee has not yet been notified of the outcome of any site-specific assessment (SSA) for the non-NHS research site(s) taking part in this study. The favourable opinion does not therefore apply to any non-NHS site at present. I will write to you again as soon as one Research Ethics Committee has notified the outcome of a SSA. In the meantime no study procedures should be initiated at non-NHS sites.

Conditions of the favourable opinion

The favourable opinion is subject to the following conditions being met prior to the start of the study.

Management permission or approval must be obtained from each host organisation prior to the start of the study at the site concerned.

This Research Ethics Committee is an advisory committee to London Strategic Health Authority.
The National Research Ethics Service (NRES) represents the NRES Directorate within
the National Patient Safety Agency and Research Ethics Committees in England

For NHS research sites only, management permission for research ("R&D approval") should be obtained from the relevant care organisation(s) in accordance with NHS research governance arrangements. Guidance on applying for NHS permission for research is available in the Integrated Research Application System or at <http://www.rdforum.nhs.uk>.

Where the only involvement of the NHS organisation is as a Participant Identification Centre (PIC), management permission for research is not required but the R&D office should be notified of the study and agree to the organisation's involvement. Guidance on procedures for PICs is available in IRAS. Further advice should be sought from the R&D office where necessary.

Sponsors are not required to notify the Committee of approvals from host organisations.

It is the responsibility of the sponsor to ensure that all the conditions are complied with before the start of the study or its initiation at a particular site (as applicable).

Approved documents

The final list of documents reviewed and approved by the Committee is as follows:

<i>Document</i>	<i>Version</i>	<i>Date</i>
Protocol	1	12 November 2010
GP/Consultant Information Sheets	1.1	17 February 2011
Response to Request for Further Information		23 February 2011
Investigator CV		
Participant Consent Form: Young people (aged 11-16 years)	1.1	17 February 2011
CV Natasha Payne		
Assent Form for children (aged 6-10 years)	1.1	17 February 2011
Participant Information Sheet	1	02 December 2010
Covering Letter		07 December 2010
Evidence of insurance or indemnity		06 December 2010
REC application		09 December 2010
Participant Information Sheet: Parent/Guardian	1	02 November 2010
Participant Information Sheet: Young people (aged 11-16 years)	1.1	17 February 2011
Participant Information Sheet: Parents/Guardians	1.1	17 February 2011
Participant Information Sheet: Adults	1.1	17 February 2011
Participant Consent Form: Adults	1.1	17 February 2011
Participant Consent Form: Parents/Guardians	1.1	17 February 2011

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees (July 2001) and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

After ethical review

Now that you have completed the application process please visit the National Research Ethics Service website > After Review

You are invited to give your view of the service that you have received from the National Research Ethics Service and the application procedure. If you wish to make your views

known please use the feedback form available on the website.

The attached document "*After ethical review – guidance for researchers*" gives detailed guidance on reporting requirements for studies with a favourable opinion, including:

- Notifying substantial amendments
- Adding new sites and investigators
- Progress and safety reports
- Notifying the end of the study

The NRES website also provides guidance on these topics, which is updated in the light of changes in reporting requirements or procedures.

We would also like to inform you that we consult regularly with stakeholders to improve our service. If you would like to join our Reference Group please email referencegroup@nres.npsa.nhs.uk.

11/H0721/4

Please quote this number on all correspondence

With the Committee's best wishes for the success of this project

Yours sincerely



Dr David Slovic
Chair

Email:

Enclosures: "After ethical review – guidance for researchers"

Copy to: Mr Dave Wilson, R&D office
Miss Natasha Payne

Appendix 2.2



Health Research Authority

NRES Committee London - City Road & Hampstead

Bristol Research Ethics Centre
Level 3, Block B
Whitefriars
Levens Mound
Bristol
BS1 2VT

Tel: 0117 342 1330
Fax: 0117 342 0445

13 December 2011

Professor Sanjay Sisodiya
UCL Institute of Neurology
National Hospital for Neurology and Neurosurgery
Queens Square
London
WC1N 3BG

Dear Professor Sisodiya

Study title: A Genetic Basis for Response to the Ketogenic Diet in Epilepsy
REC reference: 11/H0721/4
Protocol number: 10/0427
Amendment number: 1
Amendment date: 05 October 2011

The above amendment was reviewed at the meeting of the Sub-Committee held on by the correspondence.

Ethical opinion

There were no ethical issues raised.

The members of the Committee taking part in the review gave a favourable ethical opinion of the amendment on the basis described in the notice of amendment form and supporting documentation.

Approved documents

The documents reviewed and approved at the meeting were:

Document	Version	Date
Notice of Substantial Amendment (non-GTAMPs)	1	05 October 2011
Protocol	1.1	05 October 2011
Participant Information Sheet: Adult	1.2	05 October 2011
Participant Consent Form: Adult	1.2	05 October 2011
Participant Information Sheet: Parent/Guardian	1.2	05 October 2011
Participant Information Sheet: Young person	1.2	05 October 2011
Participant Information Sheet: Child	1.1	05 October 2011
Participant Consent Form: Parent/Guardian	1.2	05 October 2011

Participant Consent Form: Young person	1.2	05 October 2011
Participant Consent Form: Child Assent form	1.2	05 October 2011

Membership of the Committee

The members of the Committee who took part in the review are listed on the attached sheet.

R&D approval

All investigators and research collaborators in the NHS should notify the R&D office for the relevant NHS care organisation of this amendment and check whether it affects R&D approval of the research.

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees (July 2001) and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

11/H0721/4:	Please quote this number on all correspondence
--------------------	---

Yours sincerely

Dr David Slovic
Chair

E-mail: David.Slovic@ucl.ac.uk

Enclosures: List of names and professions of members who took part in the review

*Copy to: Mr Dave Wilson
Mr Philip Diamond, University College London Hospitals NHS
Foundation Trust*

Appendix 2.3

The National Hospital for Neurology & Neurosurgery
Queen Square
London
WC1N 3BG

Telephone: 0845 155 5000
Web-site: www.uclh.nhs.uk

ADULT PATIENT INFORMATION SHEET **Genetics of Response to Ketogenic Diet**

You are invited to participate in a study that is being carried out at the National Hospital for Neurology and Neurosurgery in collaboration with Great Ormond Street Hospital, the National Centre for Young People with Epilepsy, Evelina Children's Hospital and Birmingham Children's Hospital. Before you decide, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully. One of our team will go through the information sheet with you and answer any questions you have. There will be the opportunity to ask questions if anything is not clear, or if you would like further information.

PART 1

What is the purpose of this study?

The aim of this study is to see whether there is a genetic basis to treatment outcome with the Ketogenic diet and whether the genetic basis might be because the diet causes changes in gene expression.

Why have I been invited?

As will have been discussed with you in clinic, we still do not understand why some people respond well to the Ketogenic diet and others do not. The purpose of this study is to collect together some patients who have received, or are currently receiving the Ketogenic diet and look at their genetic material to see whether there are any similarities in those who have responded well to treatment and those who have not. We will also look at whether the Ketogenic diet causes changes in patients' genes. By looking at a group of people treated with the diet, we hope that we will find some common genetic changes and improve our understanding of how the Ketogenic diet works.

Do I have to take part?

You do not have to take part in the study if you do not wish to. If you decide to take part, you may withdraw your information at any time without having to give a reason. The decision about whether to take part or not will not affect your care and management in any way. If you decide to take part, you will be given a copy of this information sheet to keep and be asked to sign a consent form, a copy of which you will also be given to keep. If you decide to take part, you are still free to withdraw your data at any time, without giving a reason.

What will happen to me if I take part?

If you have not yet started the diet, we will take two or three blood samples from you. The first blood sample will be taken before you start the diet; the second sample will be taken at one point while you are on the diet. If you are weaned off the diet while the study is still in operation, a third blood sample will be taken.

If you are already on the diet, we will take one or two blood samples from you. The first blood sample will be taken at one point while you are on the diet. If you are weaned off the diet while the study is still in operation, a second blood sample will be taken.

If you are no longer following the diet, we will take a single blood sample from you.

On all occasions, it is likely that we will take the blood at the same time as blood taken for routine clinical monitoring (this means that an extra blood sample will be taken). We will require a very small amount of blood – 10mls. The procedure will be like any previous blood sample that has been undertaken. The blood will be taken by an experienced blood taker. The procedure should not cause any problems, but very occasionally it can result in mild bruising or discomfort. The blood sample will be sent to the laboratory to isolate genetic material (DNA and RNA) from the blood. Tests will then be performed when we have a collection of samples to see if there are any changes in the genes that are common to people with similar types of epilepsy. We will not be able to return any individual results. However, at the end of the study we will make the overall results available to all those who participated.

Seizure frequency data will be obtained from that already collected as part of the monitoring for the KD service provision.

All the information we gather will be held on a computer at the Institute of Neurology. Only Professor Sisodiya, his collaborator at Great Ormond Street Hospital, Professor Cross, and the research team undertaking this study will have access to it. Professor Sisodiya and Professor Cross will have sole responsibility for access to this information. At the end of the study the sample will be kept for 20 years for future ethically approved projects. The nature of any further research will depend on our initial findings. The information and results we collect will remain confidential forever to uphold good research practice. The information you give will only be used for medical research and will be stored and disposed of securely.

What will I have to do?

As outlined above, this research involves taking one, two or three blood tests from which genetic material will be extracted and analysed. No extra visits to the hospital will be required.

What are the possible disadvantages and risks of taking part?

Taking part in this research will not affect the management of your epilepsy. The only risk is in minor bruising from the blood test. You may ask for all information collected for this study to be destroyed at any time. All information regarding your medical records will be treated as strictly confidential and will only be used for medical and research purposes. Your medical records may be inspected by constant authorities and properly authorised persons, but if any information is released this will be done so in a coded form so that confidentiality is strictly maintained.

What are the possible benefits of taking part?

The results of this research are unlikely to help you specifically. However, in the long term, it may influence treatment choices in individuals with a similar epilepsy.

What if there is a problem?

Any complaints about the way you have been dealt with during the study or any possible harm you might suffer will be addressed. The detailed information concerning this is given in Part 2 of this information sheet.

Will my taking part in the study be kept confidential?

All information about the study will be kept strictly confidential; we will follow ethical and legal practice. The details are included in part 2.

PART 2 INFORMATION YOU NEED TO KNOW IF YOU WANT TO TAKE PART

What will happen if I don't want to carry on with the study?

If you wish to withdraw from the study at any time, we will destroy all identifiable samples, but we will need to use the data collected up to the time of your withdrawal.

What if there is a problem?

Every care will be taken in the course of this study. However, in the unlikely event that you are injured by taking part, compensation may be available.

If you suspect that the injury is the result of the Sponsor's (University College London) or the hospital's negligence then you may be able to claim compensation. After discussing with your research doctor, please make the claim in writing to Professor Sanjay Sisodiya, who is the Chief Investigator for the research and is based at the Institute of Neurology, 33 Queen Square, London WC1N 3BG; telephone number: 020 3448 8612; fax number: 020 3448 8615; email address: s.sisodiya@ucl.ac.uk. The Chief Investigator will then pass the claim to the Sponsor's Insurers, via the Sponsor's office. You may have to bear the costs of the legal action initially, and you should consult a lawyer about this.

Regardless of this, if you wish to complain, or have any concerns about any aspect of the way you have been approached or treated by members of staff or about any side effects (adverse events) you may have experienced due to your participation in the research, the normal National Health Service complaints mechanisms are available to you. Please ask your research doctor if you would like more information on this. Details can also be obtained from the Department of Health website:

<http://www.dh.gov.uk>.

Will my taking part in this study be kept confidential?

This information will be held on a computer at the Institute of Neurology and only Professor Sisodiya, his collaborator at the Institute of Child Health, Professor Cross, and the research team undertaking this study will have access to it. Professor Sisodiya and Professor Cross will have sole responsibility for access to this information. At the end of the study, the information will be kept for further research. The nature of which will depend on our initial findings.

In line with current practice in the field of genetics, we plan to make overall results which cannot be related back to individual people, openly available upon publication of the main findings from the study. There is no way anybody will be able to identify you from the study. Coded individual genetic results, which also cannot be traced

back to individual people, will be made available only to qualified researchers with strict agreements on use of the information in due course. This is to further enhance the value of the research and to maximise the potential of the findings for research and epilepsy.

Involvement of the General Practitioner

Your GP will be informed of your participation in the study. No information will be released without your consent.

What will happen to any samples taken from me?

One, two or three blood samples will be taken. These will go to the laboratory where genetic material (DNA and RNA) will be extracted. Information will be held on a computer at the Institute of Neurology and only Professor Sisodiya, his collaborator at the Institute of Child Health, Professor Cross, and the research team undertaking this study will have access to it. Professor Sisodiya and Professor Cross will have sole responsibility for access to this information. At the end of the study the sample will be kept for further research. Samples will be stored for 20 years. The nature of any further research will depend on our initial findings. They will not be used for other research without your consent.

In the unlikely event of a loss of capacity, the research team would retain blood and personal data collected and continue to use it confidentially in connection with the purposes for which consent is being sought.

What will happen to the results of the research study?

On the whole, results will not be available on an individual basis. In some cases a significant genetic result may be found. If this is the case, we will arrange to discuss this with you and your consultant, and arrange further investigations as necessary. The overall results at the end of the study will be made available to all participants. Results will be published in medical journals. No patient will be identified in any report or publication unless consent has been obtained.

Who is organising and funding the research?

The research is being organised through UCL-Institute of Neurology in collaboration with Great Ormond Street Hospital for Children NHS Trust. It is being funded by The Epilepsy Society and UCL.

Who has reviewed the study?

All research in the NHS is looked at by an independent group of people called a Research Ethics Committee, to protect you and your child's interests. This study has been reviewed and given a favourable opinion by East Central London Research Ethics Committee 1.

Further information and contact details

1. For general information about research please review the UKCRN website: www.UKCRN.org.uk
2. If you require specific information about this research project, or are unhappy with the study, please contact Professor Sanjay Sisodiya at the Institute of Neurology, 33 Queen Square, London WC1N 3BG; telephone number: 020 3448 8612; fax number: 020 3448 8615; email address: s.sisodiya@ucl.ac.uk
3. For advice about whether to participate please consult with your consultant at the National Hospital for Neurology and Neurosurgery.

The National Hospital for Neurology & Neurosurgery
Queen Square
London
WC1N 3BG

Telephone: 0845 155 5000
Web-site: www.uclh.nhs.uk

ADULT PATIENT CONSENT FORM

Title of the Research project: Genetics of Response to Ketogenic Diet

Study Protocol No: 10/0427

**Researchers: Professor Sanjay M. Sisodiya, Professor Josemir Sander,
Professor J. Helen Cross, Miss Natasha Payne**

Contact details: ☎ 020 3448 8612: 💻 email: s.sisodiya@ucl.ac.uk

Patient Identification No for this trial:

Please initial box to indicate agreement:

1	I confirm that I have read and understand the information sheet for the above study (Version 1.2 dated 05/10/11). I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.	
2	I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, without my medical care or legal rights being affected.	
3	I understand that relevant sections of any of my Medical Notes and data collected during the study may be looked at by employees from Regulatory Authorities or from National Hospital for Neurology and Neurosurgery, Great Ormond Street Hospital/ Institute of Child Health, National Centre for Young People with Epilepsy, Evelina Children's Hospital or Birmingham Children's Hospital, where it is relevant to my taking part in this research. I give permission for these individuals to have access to my records.	
4	I agree to my GP and/or neurologist being informed of my participation in the study.	
5	I agree for my blood and DNA/RNA sample to be taken once, twice or three times at a routine clinic visit and stored with my medical records for the purpose of this study.	
6	I agree for my blood and DNA/RNA sample to be stored for future ethically approved research.	
7	I agree to take part in the above study.	

Name of Patient

Date

Signature

Name of Person taking consent
(if different from Investigator)

Date

Signature

Investigator

Date

Signature

PARENT/GUARDIAN INFORMATION SHEET

Genetics of Response to Ketogenic Diet

Your child is invited to participate in a study that is being carried out at Great Ormond Street Hospital in collaboration with the National Hospital for Neurology and Neurosurgery, The National Centre for Young People with Epilepsy, Evelina Children's Hospital and Birmingham Children's Hospital. Before you decide, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully. One of our team will go through the information sheet with you and answer any questions you have. There will be the opportunity to ask questions if anything is not clear, or if you would like further information.

PART 1

What is the purpose of this study?

The aim of this study is to see whether there is a genetic basis to treatment outcome with the Ketogenic diet and whether the genetic basis might be because the diet causes changes in gene expression.

Why has my child been invited?

As will have been discussed with you and your child in clinic, we still do not understand why some people respond well to the Ketogenic diet and others do not. The purpose of this study is to collect together some patients who have received, or are currently receiving the Ketogenic diet and look at their genetic material to see whether there are any similarities in those who have responded well to treatment and those who have not. We will also look at whether the Ketogenic diet causes changes in patients' genes. By looking at a group of people treated with the diet, we hope that we will find some common genetic changes and improve our understanding of how the Ketogenic diet works

Does my child have to take part?

Your child does not have to take part in the study if he or she does not wish to. We are also releasing information to your child for them to also make the decision. If your child decides to take part you may withdraw your child's information at any time without having to give a reason. The decision about whether to take part or not will not affect your child's care and management in any way. If your child decides to take part you will be given a copy of this information sheet to keep and be asked to sign a consent form, a copy of which you will also be given to keep. We also have an assent form for your child to sign if he or she so wishes. If your child decides to take part you are still free to withdraw your child's data at any time and without giving a reason

What will happen to my child if he or she takes part?

If your child has not yet started the diet, we will take two or three blood samples from your child. The first blood sample will be taken before your child starts the diet; the second sample will be taken at one point while he/she is on the diet. If your child is weaned off the diet while the study is still in operation, a third blood sample will be taken.

If your child is already on the diet, we will take one or two blood samples from your child. The first blood sample will be taken at one point while your child is on the diet. If your child is weaned off the diet while the study is still in operation, a second blood sample will be taken.

If your child is no longer following the diet, we will take a single blood sample from your child.

We will require a very small amount of blood: 2-5mls. On all occasions, it is likely that we will take the blood at the same time as blood taken for routine clinical monitoring (this means that an extra blood sample will be taken). The procedure will be like any previous blood sample that has been undertaken in your child. We will use anaesthetic spray if you and your child so wish, and the blood will be taken by an experienced blood taker. The procedure should not cause any problems, but very occasionally it can result in mild bruising or discomfort. The blood sample will be sent to the laboratory to isolate genetic material (DNA and RNA) from the blood. Tests will then be performed when we have a collection of samples to see if there are any changes in the genes that are common to children with similar types of epilepsy. We will not be able to return any individual results, however at the end of the study we will make the overall results available to all those who participated. Seizure frequency data will be obtained from that already collected as part of the monitoring for the KD service provision.

All the information we gather will be held on a computer at the Institute of Neurology and only Professor Cross, her collaborator at the Institute of Neurology, Professor Sisodiya, and the research team undertaking this study will have access to it. Professor Cross and Professor Sisodiya will have sole responsibility for access to this information. At the end of the study the sample will be kept for 20 years for future ethically approved projects. The nature of any further research will depend on our initial findings. The information and results we collect will remain confidential forever to uphold good research practice. The information you give will only be used for medical research and will be stored and disposed of securely.

What will my child have to do?

As outlined above, this research involves taking one, two or three blood tests from which genetic material will be extracted and analysed. No extra visits to the hospital will be required.

What are the possible disadvantages and risks of taking part?

Taking part in this research will not affect the management of your child's epilepsy. The only risk is in minor bruising to your child from the blood test. You may ask for all information collected for this study to be destroyed at any time. All information

regarding your child's medical records will be treated as strictly confidential and will only be used for medical and research purposes. Your child's medical records may be inspected by constant authorities and properly authorised persons, but if any information is released this will be done so in a coded form so that confidentiality is strictly maintained.

What are the possible benefits of taking part?

The results of this research are unlikely to help your child specifically. However, in the long term, it may influence treatment choices in individuals with a similar epilepsy.

What if there is a problem?

Any complaint about the way your child has been dealt with during the study or any possible harm he or she might suffer will be addressed. The detailed information concerning this is given in Part 2 of this information sheet.

Will my child's taking part in the study be kept confidential?

All information about the study will be kept strictly confidential; we will follow ethical and legal practice. The details are included in part 2.

PART 2 INFORMATION YOU NEED TO KNOW IF YOU WANT TO TAKE PART

What will happen if my child doesn't want to carry on with the study?

If your child wishes to withdraw from the study at any time, we will destroy all identifiable samples, but we will need to use the data collected up to the time of your withdrawal.

What if there is a problem?

Every care will be taken in the course of this study. However, in the unlikely event that you are injured by taking part, compensation may be available.

If you suspect that the injury is the result of the Sponsor's (University College London) or the hospital's negligence then you may be able to claim compensation. After discussing with your research doctor, please make the claim in writing to Professor Sanjay Sisodiya, who is the Chief Investigator for the research and is based at the Institute of Neurology, 33 Queen Square, London WC1N 3BG; telephone number: 020 3448 8612; fax number: 020 3448 8615; email address: s.sisodiya@ucl.ac.uk. The Chief Investigator will then pass the claim to the Sponsor's Insurers, via the Sponsor's office. You may have to bear the costs of the legal action initially, and you should consult a lawyer about this.

Regardless of this, if you wish to complain, or have any concerns about any aspect of the way you or your child have been approached or treated by members of staff or about any side effects (adverse events) you may have experienced due to your participation in the research, the normal National Health Service complaints mechanisms are available to you. Please ask your research doctor if you would like more information on this. Details can also be obtained from the Department of Health website: <http://www.dh.gov.uk>.

Will my child's taking part in this study be kept confidential?

This information will be held on a computer at the Institute of Neurology and only Professor Cross, her collaborator at the Institute of Neurology, Professor Sisodiya, and the research team undertaking this study will have access to it. Professor Cross and Professor Sisodiya, will have sole responsibility for access to this information. At the end of the study, the information will be kept for further research. The nature of which will depend on our initial findings.

In line with current practice in the field of genetics we plan to make overall results which cannot be related back to individual people, openly available upon publication of the main findings from the study. There is no way anybody will be able to identify your child from the study. Coded individual genetic results which also cannot be traced back to the individual people will be made available only to qualified researchers with strict agreements on use of the information in due course. This is to further enhance the value of the research and to maximise the potential of the findings for research and epilepsy.

Involvement of the General Practitioner

Your child's GP will be informed of your child's participation in the study. No information will be released without your consent.

What will happen to any samples from my child?

Either one, two or three blood samples will be taken. These will go to the laboratory where genetic material (DNA and RNA) will be extracted. Information will be held on a computer at the Institute of Neurology and only Professor Cross, her collaborator at the Institute of Neurology, Professor Sisodiya, and the research team undertaking this study will have access to it. Professor Cross and Professor Sisodiya, will have sole responsibility for access to this information. At the end of the study the sample will be kept for further research. Samples will be stored for 20 years. The nature of any further research will depend on our initial findings. They will not be used for other research without your consent.

In the unlikely event of a loss of capacity, the research team would retain blood and personal data collected and continue to use it confidentially in connection with the purposes for which consent is being sought.

What will happen to the results of the research study?

On the whole, results will not be available on an individual basis. In some cases a significant genetic result may be found. If this is the case in your child, we will arrange to discuss this with you and your consultant, and arrange further investigations as necessary. The overall results at the end of the study will be made available to all participants. Results will be published in medical journals. No patient will be identified in any report or publication unless consent has been obtained.

Who is organising and funding the research?

The research is being organised through Great Ormond Street Hospital for Children NHS Trust in collaboration with the UCL- Institute of Neurology. It is being funded by The Epilepsy Society and UCL.

Who has reviewed the study?

All research in the NHS is looked at by an independent group of people called a Research Ethics Committee, to protect you and your child's interests. This study has been reviewed and given a favourable opinion by East Central London Research Ethics Committee 1.

Further information and contact details

1. For general information about research please review the UKCRN website: www.UKCRN.org.uk
2. If you require specific information about this research project, or are unhappy with the study, please contact Professor Helen Cross, c/o Mrs Anne Brown, Neurosciences Unit, UCL Institute of Child Health, 4/5 Long Yard, London WC1N 3LU; Telephone number: 0207 599 4105; email a.brown@ucl.ac.uk
3. For advice about whether to participate please consult with your child's consultant at Great Ormond Street Hospital for Children

INFORMATION SHEET for young people age 11-16 years

Genetics of Response to Ketogenic Diet

You are being asked to take part in a research study. This is because you have been treated with the Ketogenic diet, and we are still trying to find out why this treatment works with some people but not others. This study is trying to find out whether a change in the genetic building blocks of your body has affected whether the Ketogenic diet works in your case. Before you decide if you want to join in, it's important to understand why the research is being done, and what it will involve for you. So please consider this leaflet carefully. Talk to your family, friends, doctor or nurse if you want to.

PART 1

Why are we doing this research?

As your doctor will have explained, we still do not understand why some people respond well to the Ketogenic diet and others do not. We want to look at the genetic makeup of lots of patients with epilepsy treated with the Ketogenic diet to see whether we can find a common link. We also want to see if the Ketogenic diet causes changes in your genes. If this research is successful, we may understand better how the Ketogenic diet works.

Why have I been invited to take part?

You have come to the epilepsy clinic at Great Ormond Street Hospital to talk about your epilepsy. You and your parents will have discussed this with the consultant. As we do not know what makes the Ketogenic diet work in some people but not in others, your consultant has suggested we use some genetic material from you (taken from your blood) for this study. We hope 100 children and other people like you will participate.

Do I have to take part?

No, you don't have to take part. It's up to you. We will ask for your assent and then ask if you would sign a form. We will give you a copy of this information sheet and your signed form to keep. You are free to stop taking part at any time during the research without giving a reason. If you decide to stop this will not affect the care you receive.

What will happen to me, and what will I be asked to do if I take part?

If you have not yet started the diet, we will take two or three blood samples from you. The first blood sample will be taken before you start the diet; the second sample will be taken at one point while you are on the diet. If you are weaned off the diet while the study is still in operation, a third blood sample will be taken.

If you are already on the diet, we will take one or two blood samples from you. The first blood sample will be taken at one point while you are on the diet. If you are

weaned off the diet while the study is still in operation, a second blood sample will be taken.

If you are no longer following the diet, we will take a single blood sample from you.

We can use anaesthetic spray or cream if you wish. The blood sample will go for further analysis in the laboratory. We will also record some details about your epilepsy from talking with you and your parents, as well as your hospital notes. All the details we collect will be recorded on a database but they will be coded so that they cannot be traced back to you. It will not affect your treatment in any way.

What are the possible benefits of taking part?

We cannot promise the study will help you but the information we get might help treat people with epilepsy in the future.

What if I don't want to do the research anymore?

If at any time you don't want to do the research anymore, just tell your parents, doctor or nurse. They will not be cross with you. We will destroy all identifiable samples, but we will need to use the data collected up to the time of your withdrawal.

Contact details

If you have any further questions please get in touch with Professor Helen Cross at the Neurosciences Unit, UCL Institute of Child Health, 4/5 Long Yard, London WC1N 3LU; Telephone number: 0207 599 4105.

PART 2 INFORMATION YOU NEED TO KNOW IF YOU WANT TO TAKE PART

What if there is a problem or something goes wrong?

If there is a problem, please get in touch with Professor Sanjay Sisodiya at the Institute of Neurology, 33 Queen Square, London WC1N 3BG; telephone number: 020 3448 8612; fax number: 020 3448 8615; email address: s.sisodiya@ucl.ac.uk. If you are unhappy then please talk with your parents; there is a way to formally complain if you want to.

Will anyone else know I'm doing this?

We will keep your information in confidence. This means we will only tell those who have a need or right to know.

What will happen to any samples I give?

One, two or three blood samples will be taken. These will go to the laboratory where genetic material (DNA and RNA) will be extracted. Information will be held on a computer at the Institute of Neurology and only Professor Cross, her collaborator at the Institute of Neurology, Professor Sisodiya, and her research team undertaking this study will have access to it. Professor Cross and Professor Sisodiya, will have sole responsibility for access to this information. At the end of the study the sample will be kept for further research. The samples will be kept for 20 years. The nature of any further research will depend on our initial findings. They will not be used for other research without your consent.

Who is organising and funding the research?

The research is being organised by Great Ormond Street Hospital for Children NHS Trust with the National Hospital for Neurology and Neurosurgery. The research is being funded by The National Society for Epilepsy and UCL.

Who has reviewed the study?

Before any research goes ahead it has to be checked by a Research Ethics Committee. They make sure that the research is fair. Your project has been checked by East Central London Research Ethics Committee 1.

Thank you for reading this. Please ask any questions if you need to.

INFORMATION SHEET FOR CHILDREN AGE 6-10 YEARS GENETICS AND THE KETOGENIC DIET

Why is the research being done?

Research is a way we try to find out the answers to questions. In this research we are trying to find out the reasons why some people get better with the Ketogenic diet and some people don't.

Why have I been asked to take part?

You have come to Great Ormond Street Hospital to talk about your epilepsy and a special diet - the Ketogenic diet. We are not sure exactly why some people get better with the Ketogenic diet and some people don't. We think we can find out more by looking at your blood and that of other people on the Ketogenic diet.

Did anyone else check the study is ok to do?

Before any research is allowed to happen, it has to be checked by a group of people called a Research Ethics Committee. They make sure that the research is fair. This project has been checked by the East Central London Research Ethics Committee 1.

Do I have to take part?

You do not have to take part if you don't want to. If you don't want to it will not affect your treatment in any way. Your doctor will still continue to see you.

What will happen to me if I take part in the research?

You will have one, two or three blood tests. We can use magic spray if you want us to.

Will joining in help me?

Joining in will not immediately help you, but may help other people in the future.

Will my medical details be kept private? Will anyone else know I'm doing this?

All your medical records will be kept completely private; we will tell your GP that you have taken part but no-one else will know.

What if I don't want to do the research anymore?

If at any time you don't want to do the research anymore, just tell your parents, doctor or nurse. They will not be cross with you. We will destroy all samples that have been taken.

PARENT/GUARDIAN CONSENT FORM

Title of the Research project: Genetics of Response to Ketogenic Diet

Study Protocol No: 10/0427

Researchers: Professor J. Helen Cross, Professor Sanjay M. Sisodiya, Professor Josemir Sander, Miss Natasha Payne

Contact details:  : 0207 599 4105  PA: a.brown@ucl.ac.uk

Patient Identification No for this trial:

Please **initial box** to indicate agreement:

1	I confirm that I have read and understand the information sheet for the above study (Version 1.2 dated 05/10/11). I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.	
2	I understand that my child's participation is voluntary and that he/she is free to withdraw at any time, without giving any reason, without his/her medical care or legal rights being affected.	
3	I understand that relevant sections of any of my child's Medical Notes and data collected during the study may be looked at by employees from Regulatory Authorities or from Great Ormond Street Hospital/ Institute of Child Health, the National Centre for Young People with Epilepsy, the National Hospital for Neurology and Neurosurgery, Evelina Children's Hospital, or Birmingham Children's Hospital where it is relevant to my child's taking part in this research. I give permission for these individuals to have access to my child's records.	
4	I agree to my child's GP, paediatrician and/or neurologist being informed of his/her participation in the study.	
5	I agree for my child's blood and DNA/RNA sample to be taken once, twice or three times at a routine clinic visit and stored with his/her medical records for the purpose of this study.	
6	I agree for my child's blood and DNA/RNA sample to be stored for future ethically approved research.	
7	I agree to my child taking part in the above study.	

Name of Child

Name of Parent/Guardian

Date

Signature

Name of Person taking consent
(if different from Investigator)

Date

Signature

Investigator

Date

Signature

YOUNG PERSON CONSENT FORM (AGE 11-16)

Title of the Research project: Genetics of Response to Ketogenic Diet

Study Protocol No: 10/0427

Researchers: Professor J. Helen Cross, Professor Sanjay M. Sisodiya, Miss Natasha Payne

Contact details: ☎ : 0207 599 4105 📧 PA: a.brown@ucl.ac.uk

Patient Identification No for this trial:

Please **initial box** to indicate agreement:

1	I confirm that I have read and understand the information sheet for the above study (Version 1.2 dated 05/10/11). I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.	
2	I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, without my medical care or legal rights being affected.	
3	I understand that relevant sections of any of my Medical Notes and data collected during the study, may be looked at, by employees from Regulatory Authorities or from Great Ormond Street Hospital/ Institute of Child Health, the National Centre for Young People with Epilepsy, the National Hospital for Neurology and Neurosurgery, Evelina Children's Hospital, or Birmingham Children's Hospital where it is relevant to my taking part in this research. I give permission for these individuals to have access to my records.	
4	I agree to my GP, paediatrician and/or neurologist being informed of my participation in the study.	
5	I agree for my blood and DNA/RNA sample to be taken once, twice or three times at a routine clinic visit and stored with my medical records for the purpose of this study.	
6	I agree for my blood and DNA/RNA sample to be stored for future ethically approved research.	
7	I agree to take part in the above study.	

Name of Patient

Date

Signature

Name of Person taking consent
(if different from Investigator)

Date

Signature

Investigator

Date

Signature

ASSENT FORM FOR CHILDREN (AGE 6-10)

Title of project: *Genetics of Response to Ketogenic Diet*

Study Protocol Number: 10/0427

Researchers: Professor J. Helen Cross, Professor Sanjay M. Sisodiya,
Professor Josemir Sander, Miss Natasha Payne

Contact details: ☎ : 0207 599 4105 💻 PA: a.brown@ucl.ac.uk

Patient Identification Number for this trial:

Child (or if unable, parent on their behalf) / young person to circle all they agree with:

Have you read (or had read to you) about this work (Version 1.1 dated 05/10/11)?	Yes/No
Has somebody else explained this work to you?	Yes/No
Do you understand what this work is about?	Yes/No
Have you asked all the questions you want?	Yes/No
Have you had your questions answered in a way you understand?	Yes/No
Do you understand it's OK to stop taking part at any time?	Yes/No
Are you happy to take part?	Yes/No

If any answers are '**No**' or you don't want to take part, don't sign your name!

If you do want to take part, you can write your name below:

Your name _____

Date _____

The doctor who explained this project to you needs to sign too:

Print Name _____

Sign _____

Date _____

Thank you for your help

Appendix 2.4

Ketogenic Dietary Therapies

People may find a way of following Ketogenic Dietary Therapies that suits them. This may differ from what their doctor or dietitian has said. Here are some ways in which people may follow Ketogenic Dietary Therapies. For each statement, **please circle which best applies to you.**

All questions refer to your prescribed Ketogenic Dietary Therapy.

I alter my diet

Never rarely sometimes often very often

I forget to follow the diet

Never rarely sometimes often very often

I stop following the diet for a while

Never rarely sometimes often very often

I decide to miss a meal

Never rarely sometimes often very often

Do you ever eat foods that are not allowed on the diet?

Never rarely sometimes often very often

I avoid following the diet if I can

Never rarely sometimes often very often

Thank you very much for your cooperation!

Ketogenic Dietary Therapies

People may find a way of following Ketogenic Dietary Therapies that suits them. This may differ from what their doctor or dietitian has said. Here are some ways in which people may follow Ketogenic Dietary Therapies. For each statement, **please circle which best applies to you and your child.**

All questions refer to your child's prescribed Ketogenic Dietary Therapy.

We alter the diet

Never rarely sometimes often very often

We forget to follow the diet

Never rarely sometimes often very often

My child stops following the diet for while

Never rarely sometimes often very often

My child misses a meal

Never rarely sometimes often very often

My child eats eat foods that are not allowed on the diet

Never rarely sometimes often very often

My child avoids following the diet if he/she can

Never rarely sometimes often very often

Thank you very much for your cooperation!

Appendix 4.1

Table 4.9: Results of association analysis, not accounting for ethnicity: common and intermediate variation in *KCNJ11* and *BAD* in Ketogenic diet overall responders and non-responders (summary KD response, n=216)

Gene	SNP rs number	Minor allele (in Ketogenic diet cohort)	Frequency of minor allele in non- responders	Frequency of minor allele in responders	Unadjusted p- value	Odds ratio [95%CI]	P-value (100,000 permutations)
<i>KCNJ11</i>	rs8175351	A	0.02857	0.004505	0.0615	2.085 [0.908-4.786]	0.3306
	rs1800467	G	0.08095	0.04054	0.1044	0.6757 [0.3911-1.167]	0.5055
	rs5219 (same for rs5215)	A	0.3143	0.3514	0.4749	1.99 [0.7225-5.481]	0.9842
	rs5218	T	0.2714	0.3063	0.4581	0.8462 [0.5666-1.264]	0.9782
	rs5216	G	0.01905	0.03604	0.3832	0.5194 [0.1541-1.751]	0.9463
<i>BAD</i>	rs34882006	A	0.05238	0.02703	0.2186	0.8437 [0.556-1.28]	0.764
	rs2286615	T	0.119	0.1667	0.1716	0.8462 [0.5666-1.264]	0.7052

Appendix 4.2

Table 4.10: Results of association analysis, including ethnicity as a covariate: common and intermediate variation in *KCNJ11* and *BAD* in Ketogenic diet overall responders and non-responders (summary KD response, n=216)

Gene	SNP rs number	Unadjusted p-value	Odds ratio [95%CI]	P-value (100,000 permutations)
<i>KCNJ11</i>	rs8175351	0.05897	6.327 [0.7137-56.09]	0.4623
	rs1800467	0.0901	2.06 [0.8865-4.788]	0.527
	rs5219	0.7765	0.9421 [0.6245-1.421]	1
	rs5218	0.4147	0.8375 [0.5475-1.281]	0.9803
	rs5216	0.3097	0.5361 [0.1585-1.814]	0.9537
<i>BAD</i>	rs34882006	0.1521	2.075 [0.7504-5.74]	0.6971
	rs2286615	0.1743	0.678 [0.388-1.185]	0.7334

Appendix 4.3

Table 4.11: Top 10 most significant results for haplotype association analysis: variation in *KCNJ11* and *BAD* in Ketogenic diet overall responders and non-responders

(summary KD response, n=216)

SNPs in haplotype	Haplotype	Frequency in non-responders	Frequency in responders	P-value
rs41282930 rs8175351 rs1800467 rs5219 rs5218 rs5216 rs5215 rs5214 rs75052600 rs59283629 rs34882006 rs2286616	CACGCCAACCGC	0.03006	0	0.01004
rs41282930 rs8175351 rs1800467 rs5219 rs5218 rs5216 rs5215 rs5214 rs75052600 rs59283629 rs34882006	CACGCCAACCG	0.02992	0	0.01005
rs41282930 rs8175351 rs1800467 rs5219 rs5218 rs5216 rs5215 rs5214 rs75052600 rs59283629	CACGCCAACC	0.02969	0	0.01027
rs112070496 rs41282930 rs8175351 rs1800467 rs5219 rs5218 rs5216 rs5215 rs5214 rs75052600 rs59283629 rs34882006	GCACGCCAACCG	0.03023	0	0.01034
rs8175351 rs1800467 rs5219 rs5218 rs5216 rs5215 rs5214 rs75052600 rs59283629 rs34882006 rs2286616	ACGCCAACCGC	0.02975	0	0.01043
rs8175351 rs1800467 rs5219 rs5218 rs5216 rs5215 rs5214 rs75052600 rs59283629 rs34882006	ACGCCAACCG	0.02961	0	0.01044
rs112070496 rs41282930 rs8175351 rs1800467 rs5219 rs5218 rs5216 rs5215 rs5214 rs75052600 rs59283629	GCACGCCAACC	0.02999	0	0.01049
rs8175351 rs1800467 rs5219 rs5218 rs5216 rs5215 rs5214 rs75052600 rs59283629	ACGCCAACC	0.0294	0	0.01065
rs41282930 rs8175351 rs1800467	CGC	0.8894	0.955	0.01074
rs112070496 rs41282930 rs8175351 rs1800467	GCGC	0.8894	0.955	0.01074

Appendix 5.1

Table 5.2: Known regions of long-range linkage disequilibrium (identified by Price et al [687])

excluded from Identity-by-State calculations

Chromosome	Start position (NCBI build 37)	End position (NCBI build 37)
1	48287980	52287979
2	86088342	101041482
2	134666268	138166268
2	183174494	190174494
3	47524996	50024996
3	83417310	86917310
5	97972100	100472101
5	128972101	131972101
5	135472101	138472101
6	25392021	33392022
6	139958307	142458307
8	7962590	11962591
8	111930824	114930824
11	87860352	90860352
12	111037280	113537280
20	32536339	35066586
3	88917310	96017310
5	44464243	50464243
6	56892041	63942041
7	55225791	66555850
8	42880843	49837447
10	36959994	43679994
11	46043424	57243424
12	33108733	41713733

Publications arising from this work

Payne, N.E., Cross, J. H., Sander, J.W., Sisodiya, S.M. et al., *The ketogenic and related diets in adolescents and adults - a review*. Epilepsia, 2011. **52**: p. 1941-8.

Schoeler, N.E., Cross, J. H., Sander, J. W., Sisodiya, S. M. et al., *Can we predict a favourable response to Ketogenic Diet Therapies for drug-resistant epilepsy?* Epilepsy Res, 2013. **106**(1-2): p. 1-16.

Schoeler, N.E., Wood, S., Aldridge, V., Sander, J. W., Cross, J. H., Sisodiya, S. M. Ketogenic dietary therapies for adults with epilepsy: feasibility and classification of response.

Accepted on 04/06/14 for publication in *Epilepsy & Behavior*. Ms. No.: EB-14-235

Abstracts/poster presentations:

30th International Epilepsy Congress, Montréal, Canada, 2013: poster presentations and publication of abstracts.

Schoeler, NE; Cross, JH; Sander, JW; Sisodiya, SM; (2013) A GENETIC BASIS FOR TREATMENT RESPONSE TO THE KETOGENIC DIET. In: EPILEPSIA. (pp. 229 - 229)

and

Coppola, A; Tostevin, A; Leu, C; Schoeler, N; Cross, HJ; Sisodiya, SM; (2013) COPY NUMBER VARIANT BURDEN IN A PEDIATRIC POPULATION WITH EPILEPSY. In: EPILEPSIA. (pp. 201 - 201).

3rd International Symposium on Dietary Therapies for Epilepsy, Chicago, USA, 2012. Won prize for poster, 'A genetic basis for response to the ketogenic diet'.

European Epilepsy Congress Stockholm 2014: poster presentation and publication of abstract entitled 'A genetic basis for response to the ketogenic diet' (to be published June 2014)